

**Control and Vulnerability: Reflections on the Nature of Human Agency and
Personhood**

A thesis submitted in fulfilment of the
requirements for the degree of

Doctor of Philosophy
in
The Department of Philosophy

Rhodes University

Sharli Anne Paphitis

October 2014

Supervisor: Prof. Pedro Alexis Tabensky

Co-Supervisor: Dr. Richard Flockemann

Abstract

Following the writings of philosophers such as Harry Frankfurt, Gary Watson, and Alfred Mele, in this thesis I defend some central claims of the self-control view of human agency. However, I not only defend, but also supplement this view in the following two ways. First, drawing on work by Mary Midgley and Sigmund Freud I advance the claim that self-control requires the experience of internal conflict between an agent's motivations and intentions. Second, drawing on insights from Simone de Beauvoir and Friedrich Nietzsche, as well as recent research in social psychology and cognitive science, I will argue in this thesis that self-control and vulnerability are inextricably intertwined with one another, and that as a result both are to be seen as constitutive of human agency. While it is the capacity for self-control that marks us out as human agents, I argue that it is also our uniquely human vulnerability which distinguishes our agency from the kind of agency which we might attribute to other potential or actual forms of sentience.

Further, while the concepts of human agency and personhood are typically conflated in the analytic tradition of philosophy, in this thesis I will show that there are good reasons for understanding these two concepts as subtly distinct from one another. The term personhood, I will argue, can fruitfully be understood in substantive rather than purely formal terms. A person, in the superlative sense, is to be understood as someone who exercises their agency well; and, as such, persons are answerable to a number of normative prescriptions. Following Midgley, Nietzsche and Martha Nussbaum, I argue against Frankfurt's normative prescription for personhood in the form of what he calls 'wholeheartedness', and offer four normative prescriptions for personhood of my own.

Acknowledgements

I am greatly indebted to Pedro Tabensky for his insightful instruction, guidance and mentorship. I would like to thank Pedro for his continued commitment to this project, and for his unfailing confidence in me through challenging times. I would also like to thank Diana Hornby not only for her valuable practical insights into the vulnerability of human agency and personhood, but for her relentless understanding, generosity and patience. I would like to extend a warm word of thanks to Abraham Olivier for useful feedback and commentary on this thesis and for going above and beyond the call of duty in all circumstances. Thanks also to Geoff Stapley for taking the time to engage with my ideas so thoughtfully. Most importantly, I would like to thank my husband, James Stapley, for supporting and encouraging me at every turn.

Finally, this thesis would not have been possible without the financial support of the Andrew Mellon Mentors Scholarship, the National Research Foundation¹, and the Erasmus Mundus Programme, for which I am most grateful.

¹ The financial assistance of the National Research Foundation (NRF) towards this research project is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and not necessarily to be attributed to the NRF.

List of Abbreviations

For the works of Friedrich Nietzsche published or prepared for publication, all references are by paragraph number (e.g. GS 270), or section and paragraph number (e.g. GM II: 2), as appropriate. The following are the abbreviations used for the works from which quotations have been taken:

Z *Thus Spoke Zarathustra*

SE *Schopenhauer as Educator*

WP *The Will to Power*

TI *Twilight of the Idols; The Anti-Christ*

GS *The Gay Science*

BG *Beyond Good and Evil ~ Prelude to a Philosophy of the Future*

BT *The Birth of Tragedy from the Spirit of Music*

GM *On the Genealogy of Morals*

Contents

Introduction	1
Part One: Human Agency, Self-Control, and Vulnerability	8
Chapter One: A Self-Control View of Human Agency	13
Self-Direction, Self-Control and Human Agency	16
Constraint, Sanity and Procedural Independence	27
A Disagreement about Our Better Judgement	37
Internal Conflict	44
Chapter Two: Vulnerability and the Difficulty of Self-Control	56
Vulnerability and Difficulty	61
Self-Control and Embodiment	66
Self-Control and Situation	71
Cognitive Limitations for Self-Control	81
Self-Control and Internal Conflict	89
Improved Self-Control	94
Contingently Necessary Requirements and Constitutive Claims	98
Part Two: Human Agency, Personhood and Normativity	104
Chapter Three: Self-Governance, Wholeheartedness and Integrity	108
Self-Governance, Identification, and Wholeheartedness	111
Externalisation and Dissociation	117
Gothic Subjectivity	125
Integrity	141
Chapter Four: Ambivalence, Affirmation, and Authenticity	149
Wholehearted Wholeheartedness and Ambivalence	152
Acknowledgement and Affirmation	169
Authenticity and the Art of Living	188
Conclusion	196
Reference List	198

“When we were children, we used to think that when we were grown-up we would no longer be vulnerable. But to grow up is to accept vulnerability... To be alive is to be vulnerable.”

Walking on Water: Reflections on Faith and Art, Madeleine L'Engle

“Oh, wretched ephemeral race, children of chance and misery.”

The Birth of Tragedy from the Spirit of Music, Friedrich Nietzsche

Introduction

The concepts of human agency and personhood have, most influentially in the analytic tradition of philosophy, been explored in the writings of Harry Frankfurt. In this thesis I will argue for and defend the central tenet of Frankfurt's account, namely, that self-control is constitutive of both human agency and personhood. However, in this thesis I will also revise and supplement his view. Drawing on accounts of human agency and personhood in both the analytic and continental traditions of philosophy, as well as research pertaining to these concepts in psychology, the biological and cognitive sciences, behavioural economics, and literature I will argue that both self-control and vulnerability are constitutive of human agency and personhood.

The basic project in this thesis, then, is to champion distinct conceptions of human agency and personhood which move beyond narrow, control focused understandings of what it means to be in command of oneself, towards more encompassing notions which incorporate the important consideration of our own human vulnerability.² To be human is to be vulnerable, and so our human vulnerability is, at least in part, constitutive of our agency and personhood.

Further, while the concepts of human agency and personhood are undoubtedly interrelated, the exact nature of this relationship has, to some extent, been neglected in recent philosophical work. In addressing this oversight, I will argue in this thesis that while human agency is required for personhood, the concepts of human agency and personhood should be understood as subtly distinct from one another. Personhood, I will argue, is to be understood as a substantive notion, and as such, I argue for four normative prescriptions for personhood.

² The setting for this task within this thesis involves bridging two very distinct philosophical traditions, as well as incorporating understandings of agency and personhood from work in psychology and science. As such, much of the work requires the careful unpacking of different conceptions of agency and personhood, along with the careful unpacking of some notably slippery terminology used inconsistently across disciplines. As far as it is possible I have tried to pay attention to the different and complex contexts of the literature upon which I draw in this thesis, and I aim to present a view which draws coherently across traditions and disciplines as far as this is possible.

Human Agency

In the first part of this thesis I will defend a self-control view of human agency, drawing on work in the analytic philosophical tradition, while supplementing this view with important insights from biology and cognitive science, psychology and attention to the complexities of actual human life. Questions about the nature, functioning and importance of self-control have been central to discussions about the nature of human agency in both the analytic tradition of philosophy, most importantly for my purposes here in the work of Harry Frankfurt, Gary Watson and Alfred Mele and, more recently in biology.³ I will distinguish in the first part of this thesis, between agency simpliciter, which I see as involving only very basic self-direction or goal directed behaviour, and human agency, which I see as involving a more complex form of self-control. Drawing this distinction allows us to understand how the nature of our own agency is distinct from the kind of agency exhibited by non-human animals along a spectrum of goal directed behaviour. The best pictures of human agency generated on a biological view, as well as in the analytic philosophical tradition, explain agency in terms of self-control, because our ability to act rather than simply be blown through the world like leaves in the wind requires the exercise of various psychological activities, such as self-observation or critical reflection, in order to achieve mastery over the self. If we are able to achieve self-control, proponents of such accounts claim, we are able to govern and control our actions and behaviour, and this is what makes us actors or human agents, as distinct from other organisms.

A major part of what makes a human agent's mental life so complex is the fact that, as humans, we not only have primary, or brute, responses to the world in the form of beliefs and desires, but we are able to assess those basic responses

³ Most recently Murphy and Brown have brought cognitive science and philosophical thoughts on agency together, their picture provides a biological basis for the kind of control focused picture of agency developed in the analytic tradition of philosophy which I will be discussing. While I will not be providing a literature review of the biological accounts of human agency which can be accessed, I take this sensible account which engages with philosophy to provide an access point from which to establish what minimally we as philosophers would need to engage with to construct our views of the biological account of human agency. This does not amount to an illicit appeal to scientific authority as the philosophical claims underpinning the view are still up for grabs in the thesis itself. What I am discussing is a particular biological view which gives us a good picture, for philosophical discussion, of what a biological picture of human agency might look like. See N. Murphy and W. Brown, *Did My Neurons Make Me Do It? - Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will*, Oxford, Oxford University Press, 2007.

through critical reflection, form opinions or make assessments of them, and, in some instances, we are able to change them. In this way, critical reflection is not merely a passive capacity but, rather, it is active in the sense that it involves a deliberative or evaluative element. When critically reflecting on our basic responses to the world, we evaluate whether or not to act on those responses by assessing them in light of our values, commitments, projects, aims and goals – what we might be inclined to call our ‘better judgment’. In so far as we have the capacity for critical reflection, we have control over whether or not we act on our most basic responses to the world. And in doing so, we control ourselves from the inside. When we speak about exercising our agency, then, what we mean is that we should have self-control in this sense. A number of robust and plausible philosophical accounts of agency have been proposed along these lines, most notably by Harry Frankfurt, Gary Watson and Alfred Mele.⁴ Philosophical accounts such as these outline the necessary structural features of the mind which must be in place in order for us to control our initial responses to the world and, thereby, allow us to perform actions which are calculated and controlled rather than being largely out of our control. Exercising self-control in this way allows us to make our own choices and decisions about the actions we take and the lives we come to live as a result, and is what we most basically refer to as human agency.

Despite the richness of the analytic philosophical literature on human agency, not enough attention there is paid to the actual business of living, where all manner of constraints threaten our ability to exercise self-control. While the

⁴ H. Frankfurt, ‘Freedom of the Will and the Concept of a Person’, *The Journal of Philosophy*, Vol. 68, No. 1, 1971, pp. 5-20. H. Frankfurt, ‘Identification and wholeheartedness’, in Frankfurt, H., *The Importance of what we Care About*, Cambridge, Cambridge University Press, 1988. H. Frankfurt, ‘The Faintest Passion’, in Frankfurt, H., *Necessity, Volition and Love*, Cambridge, Cambridge University Press, 1999. H. Frankfurt, ‘Autonomy, Necessity, and Love’, in Frankfurt, H., *Necessity, Volition and Love*, Cambridge, Cambridge University Press, 1999. H. Frankfurt, ‘Taking Ourselves Seriously’, in Satz, D. (ed.), *Taking Ourselves Seriously & Getting It Right*, California, Stanford University Press, 2006. G. Watson, ‘Free Agency’, in Christman, J. (Ed.), *The Inner Citadel: Essays on Individual Autonomy*, Oxford, Oxford University Press, 1989. G. Watson, *Agency and Answerability: Selected Essays*, New York, Oxford University Press, 2004. A. Mele *Irrationality: An Essay on Akrasia, Self-deception and Self-control*, Oxford, Oxford University Press, 1987. A. Mele, *Autonomous Agents: From Self-Control to Autonomy*, New York, Oxford University Press, 1995. A. Mele, *Motivation and Agency*, Oxford, Oxford University Press, 2003. A. Mele, *Backsliding: Understanding Weakness of Will*, Oxford, Oxford University Press, 2012. A. Mele, ‘Autonomy and neuroscience’, in Radoilska, L. (ed.), *Autonomy and Mental Disorder*, Oxford, Oxford University Press, 2012.

formal analytic accounts found in this literature might provide us with a compelling characterisation of human agency, they tend to ignore our lived experience of agency in all its complexity. This is done perhaps intentionally, so as to present us with a purely formal picture of human agency, but I suggest in this thesis that this formal picture requires supplementation. Following the writings of philosophers and psychologists, such as Nietzsche and Freud, who have shown ‘that there are deep currents of meaning, often crosscurrents, running through the human soul which can at best be glimpsed through a glass darkly’⁵, I will argue that not only our ability to formulate our ‘better judgment’ but also to act in accordance with it is far more complex than suggested in current analytic literature. Through the development of the broadly psychoanalytic picture particularly, what emerges is a far more complex picture of our psyche which is largely characterised by conflict and is far from transparent (even, or perhaps especially, to itself).

Our mental lives are characterised by disparate and often conflicting desires, beliefs, goals and values, making our psyche, I will argue, unmistakably divided. Drawing on fundamental insights from existentialism and recent work in the sciences, I will show that human agency can be threatened by such constraints, but also, and crucially, that human agency only emerges in contexts where it is threatened. As I see it, human agency is simultaneously (and, at first glance, perhaps paradoxically) both threatened by and necessarily dependent upon the effects of such constraints on the individual. And it is in light of this that I see the self-control view of agency requiring supplementation. I will argue in the first part of this thesis that human agency must be understood in terms of both self-control and vulnerability. Further, I suggest that vulnerability is not merely a contingent requirement for human agency. Rather, both self-control and vulnerability are constitutive of our agency. It is our vulnerability, I will argue, which not only distinguishes our agency from that of supernatural, disembodied or immortal beings, but it is also our vulnerability which positively shapes, flavours and enriches our understanding of our own agency.

⁵ J. Lear, *Open Minded: Working Out the Logic of the Soul*, Cambridge, Massachusetts, Harvard University Press, 1998, p. 28.

Personhood

In Frankfurt's paper 'Freedom of the Will and the Concept of a Person,' as well as in his later writings, there is a lack of clear distinction between the notions of human agency and personhood: he neglects, I think, to fully expound the nature of the relationship between these two concepts. This explanatory blind spot is not isolated to the Frankfurtian account, but is systemic throughout the existing literature on personal autonomy, agency, subjectivity, and freedom of the will. The second part of this thesis aims, in small part, to address this explanatory blind spot. I suggest that an account of human agency – as it is understood on the self-control view – on its own, does not give us a sufficiently rich account of personhood.

Typically, personhood is understood as encompassing those capacities of self-awareness and self-control associated with human agency, as well as other capacities, attributes and features such as: understanding ourselves as temporally situated⁶ (having notions of ourselves as situated between a past and a future) and, relatedly, having a narrative structure to our lives⁷; having moral sensibilities (being able to draw distinctions between 'right'/'appropriate' behaviour and 'wrong'/'inappropriate' behaviour); and being members of a moral community (possessing certain rights and duties; recognising and treating other persons in appropriate ways – as worthy of respect or loss of respect; and understanding ourselves in cultural and interpersonal contexts).⁸

In the second part of this thesis I will argue that an important difference between our understandings of human agency and personhood is that personhood is to be understood as a substantive notion. In order to understand

⁶ Philosophers working in the continental tradition of philosophy have focused on these issues most notably Sartre and Heidegger.

⁷ J. Velleman, *Self to self: Selected Essays*, Cambridge, Cambridge University Press, 2006.

⁸ C. Taylor and proponents of communitarian views or proponents of African Philosophy; as well as various contemporary continental philosophers who argue for the importance of intersubjectivity in our understanding of the constitution of the self such as Lacan and Derrida. C. Taylor, *Sources of the Self: The Making of the Modern Identity*, Cambridge, Cambridge University Press, 1989. A. Elliott, *Psychoanalytic Theory: An Introduction*, 2nd edn., Hampshire, Palgrave, 2002.

what our personhood consists in, then, the description of human agency provided by the self-control view must be substantiated in normative terms.

On his account of personhood, Frankfurt offers a normative prescription in the form of wholeheartedness, which for him entails the process which he calls dissociation. Contra Frankfurt, I argue that wholeheartedness is a problematic candidate for a normative prescription for personhood because it is ultimately damaging to the self. In the second part of this thesis I offer an alternative candidate for a normative prescription for personhood. Following Midgley and Freud, I see division and conflict as necessary parts of our psychology and our experience as human subjects. This division and conflict constitutes what I see as the essentially fragile internal situation in which we find ourselves as persons, and the struggle to exercise our agency well emerges in precisely this context. Drawing on examples of literary doubles given in the Gothic psychological horrors of the late Nineteenth century, I argue that to be a person is a struggle to achieve a measure of psychological integrity. Late Gothic fiction, I will argue, provides us with horror stories generated by the uncanny transcriptions of a fragile human condition which we often struggle to fully comprehend or embrace.⁹ What emerges from this picture, I will argue, provides us with a strong case against Frankfortian wholeheartedness through the process of dissociation, and in favour of integrity. Persons, I will argue, ought to cultivate the kind of integrity Frankfortian dissociation would stifle.

Drawing on the work of Friedrich Nietzsche and Martha Nussbaum¹⁰, in second part of my thesis I will argue for three further normative prescriptions for personhood.

First, I will argue for a normative prescription of ambivalence. Drawing on Nussbaum and her discussion of a case of tragic conflict in Aeschylus' Agamemnon, I will argue that it is a normative requirement for personhood that persons ought to experience a degree of ambivalence.

⁹ R. Jackson, *Fantasy: The Literature of Subversion*, London, Methuen, 1981, pp. 17-24.

¹⁰ M. Nussbaum, 'Pity and Mercy: Nietzsche's Stoicism', in Schacht R. (ed.) *Nietzsche, Genealogy, Morality*, Berkeley, Calif. University of California Press, 1994, pp.139-167. M. Nussbaum, *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*, 2nd edn., Cambridge, New York, Cambridge University Press, 2001.

Second, drawing on Nietzsche's characterisation of the Sovereign Individual and his suggestion of the Doctrine of Eternal Recurrence, I will argue for an epistemic normative prescription for personhood in the form of affirmation. I will argue that a key aspect of living well is our own recognition of, and, importantly, affirmation of, our vulnerability in the face of those aspects of our lives over which we cannot ever have control.

Third, and finally, I will argue for a normative prescription for personhood of authenticity. I will argue that authenticity is an important addition to an account of personhood, since a vital part of living well is to live authentically. In living authentically, as persons, allows us to exercise our agency well, or to live well, in a distinct and important way, namely, by living authentically, we live our lives beautifully.

Part One: Human Agency, Self-Control, and Vulnerability

What does it mean to say that I am a human agent? To say that I am a human agent is not merely to say that I am a human being. A human being need not necessarily be an agent, or at least not a full blown one. For example, children or mentally handicapped people might not be considered capable of exercising human agency, but are still considered human beings worthy of dignity and respect. Similarly, to say that I am a human agent is not simply to say that I am capable of exercising agency broadly understood. Agency, as it is broadly understood, refers to the ability of something or someone to exert an effect on or to produce some kind of change in the world. To say that I am a human agent, however, is to say that I am capable of exercising a particular and distinctive kind of agency. Human agency, I will argue in the first part of this thesis, is to be understood as distinct from the kind of agency we might attribute other actual or potential forms of sentience, because human agency uniquely requires both self-control and vulnerability.

It might be seen as controversial by some philosophers to say that we could ascribe agency to anything other than human beings, or perhaps gods. The term agency, philosophers might argue, simply describes the uniquely human, or godly, ability to perform intentional actions, or to exercise self-direction. However, recent work in biology suggests that a degree of self-direction in the form of goal directed behaviour is something that we arguably share with other living things along a gradient ranging from basic to more complex levels.¹¹ On this view, human agency can be distinguished from the kind of agency we might attribute to other manifestations of self-direction along a gradient on which the more self-direction a creature exhibits, the greater degree of agency we would attribute to it. Importantly, on this view, while non-human animals may have the ability to perform goal directed behaviours and thus be considered self-directing agents of sorts, human agency requires the most complex form of goal directed behaviour because of the complex mental systems found in human beings when compared with other animals. One advantage of adopting this view is that it allows us to describe a unique human capacity for self-direction, while leaving room in an account of agency for non-human animals. Although the debate about whether

¹¹ See N. Murphy and W. Brown, *Did My Neurons Make Me Do It?*.

non-human animals do in fact have the necessary mental attributes required for self-direction at all rages on in philosophy, I take the arguments made recently in biology to be convincing based on my own interactions with animals.¹² Further, viewing animals as agents of sorts, as such a biological picture suggests we might, should, I think, serve as a reminder that we share something important in common with other living beings. This could have implications not only for how we understand our own more 'animalistic' sides, but perhaps, even more importantly, for animal rights activism. Finally, by distinguishing agency from human agency we can arrive at a richer notion of human agency which highlights what is distinctive about human beings in psychological terms.

Following this biological insight the first chapter of this thesis defends the view that human agency, as opposed to the kind of agency we might attribute to other living things, requires the psychological ability to exercise something like freedom of the will, as Harry Frankfurt has called it, or something like free agency as Gary Watson has called it, because of the complex mental systems found in human beings. Human agency is distinctive, I argue, because the kind of self-direction exhibited by human beings amounts to something like personal autonomy, or self-control, understood in broadly psychological terms. Freedom of the will, free agency, and autonomy are terms which are often used interchangeably in the literature, and the usage of these terms differs considerably not only across disciplines, but also within the analytic tradition of philosophy itself. What I present here, in terms of the concept of human agency, is an attempt to consolidate the meaning behind the use of these different terms as they are employed. The common thread I trace here lies in our understandings of self-direction and self-control.

Human agents are typically thought of as having a very particular kind of control over their choices and actions. What is distinctive about human agents is that

¹² The sceptic, who thinks that this biological picture does not provide evidence that we should attribute agency to anything other than human beings, might be disappointed that I do not provide any definitive arguments in my thesis in favour of attributing agency to non-human animals. However, such arguments would lead me astray from the central claims presented in this thesis. And, importantly, my positive account of human agency and personhood does not ultimately hang on whether or not the sceptic accepts the distinction between agency and human agency suggested by the biological view.

they not only form intentions to act in light of their goals, but they regard their own judgements about which courses of action to follow as an ultimate, and authoritative, motivational force. Rather than being the passive subjects of constraining forces, such as in situations of coercion or manipulation, human agents are free to the extent that their actions and behaviours can be explained by reasons and motivations which are, in the relevant sense, their own. By 'relevant sense' here I mean the ability to exercise a particular kind of self-control. Drawing on the writings of Harry Frankfurt, Gary Watson and Alfred Mele in particular, I defend in the first chapter what has been presented as the core of the self-control view in seminal philosophical debates. On such a view, a human agent is to be understood as someone who can muster the motivational strength to perform actions which are, by and large, informed by what might be called her 'better judgment'. Further, on such a view we can only be considered to exercise our agency if we are suitably free from constraints which affect either our ability to act or our ability to exercise critical reflection, that is, to form our better judgments or plan our intentions. Finally, people whose better judgements or intentional actions are determined solely by third parties are to be seen, on such a view, as lacking the kind of critical or evaluative freedom necessary to exercise human agency.

Building on and supplementing this account, I will argue in the first chapter that a human agent is someone who experiences internal conflict between her motivations and intentions. As such, I will argue that self-control is to be understood as the ability to muster the motivational strength to act at a specific time in the face of competing motivations to perform other possible actions at that same time. A capacity for self-control, I suggest, requires that we have, in general, internal conflicts. While there may be particular cases in which we are not conflicted and we might still want to say that we retain self-control, understood as a general capacity, the kind of self-control necessary for human agency cannot be had without at least the potential for internal conflict. It is my view here that, just as surely as too much internal conflict destroys our capacity for self-control – deep, unyielding and constant internal conflict produces a mental life in which self-control is no longer attainable – a psyche in which there is no potential for internal conflicts similarly precludes the possibility of self-

control. In this way, internal conflict marks human agents out as distinct from beings, perhaps like God, whose mental life does not resemble ours because there is no potential for internal conflict, and whose agency is therefore not constituted by self-control.

Further, I will argue in the first chapter of this thesis that being a human agent does not require that we exercise self-control at all times, or in all aspects of our lives. This is important primarily because it helps us to distinguish between our having the capacity for human agency and our acting in a way that manifests that agency. While there has been a lack of clarity in recent debates surrounding this distinction, I defend a view of self-control which describes a general capacity which human agents necessarily must have. Just as a surgeon who is not now practicing surgery or who only practices surgery from time to time, remains a surgeon nevertheless, a human agent retains their agency even if they are not acting in a way that manifests that agency at all times. However, there are some logical limits to this: we would be hard-pressed to call someone a surgeon who had never studied surgery and who had never performed a single surgery in her life, and it would be bizarre, I think, to call someone a human agent who had never acted in a way that manifested their agency at all.

In the second chapter, I will argue that when it comes to real life, current analytic pictures of agency provide us with little insight into the practical difficulties in exercising the kind of self-control that these accounts see as necessary for human agency. While proponents of these accounts will sometimes acknowledge the possibility that self-control might be difficult to achieve under certain conditions, the extent and importance of these practical considerations are by and large not given enough robust philosophical consideration. The examples of external constraints most frequently cited by philosophers such as Frankfurt and Watson have only exposed a fraction of the kinds of constraints which human agents actually face on a daily basis, and my discussion in the second chapter of this thesis outlines the ways in which we are vulnerable to such constraints and thus seeks, to some extent, to address this explanatory gap.

Drawing on fundamental insights in the writings of Friedrich Nietzsche about the vulnerable situation in which human agents necessarily find

themselves, I will argue that human agency requires vulnerability: that vulnerability is constitutive of human agency. I will argue that because of the difficulty we face in achieving self-control, our agency is necessarily vulnerable. Our ability to exercise self-control is easily undermined, and the constant willpower that it is necessary for us to assert in order to fight against corrupting forces at every turn is ultimately subject to dissolution. It is this necessary requirement of vulnerability which I suggest has not been adequately explored by self-control views of human agency found in the kind of philosophical accounts provided by philosophers such as Frankfurt and Watson. Following the insights of Nietzsche and the existentialists (particularly Simone de Beauvoir), I supplement the analytic self-control view of human agency, as presented in the first chapter, by discussing recent work in experimental psychology, cognitive science and behavioural economics. By combining insights from existential philosophers and recent work in the sciences I advance the view that human agency is difficult to achieve on account of our vulnerability. Importantly, I suggest that agency is not only threatened by such constraints, but that our agency only emerges in contexts in which it is threatened. Our vulnerability plays a crucial role in determining the conditions for self-control to emerge in the first place. Supplementing the self-control view of human agency in this way can help us delineate the boundaries between human agency and other, perhaps similar, kinds of agency which do not require vulnerability and which would not be considered difficult to achieve – perhaps the agency of disembodied paradisaical beings, for which self-control is not necessary. I suggest, then, that the essential fragility of the human situation must be accounted for in our understanding of the distinct kind of agency exhibited by human beings.

Chapter One: A Self-Control View of Human Agency

I wish my life and decisions to depend on myself, not on external forces of whatever kind. I wish to be the instrument of my own, not of other men's acts of will. I wish to be a subject, not an object; to be moved by reasons, by conscious purposes, which are my own, not by causes which affect me, as it were, from outside.¹³

In the above quotation, Isaiah Berlin expresses the familiar human desire for freedom as self-direction. Our ability to govern and control our own actions and behaviour is widely thought to be at least one of the distinguishing marks of our humanity, for this is in part what makes us – in Berlin's terms – subjects rather than objects. We are human subjects to the extent that we are able to control and guide our own actions (exercise self-control), rather than being controlled by external forces. And this gives us a kind of freedom that is thought to make us the authors of ourselves and our lives. Being able to make our own choices and decisions about the actions we take and the lives we come to live as a result is what we most basically refer to as human agency. Typically, we believe we have the capacity for agency and behave accordingly. Despite the multiplicity of constraints that we face in our daily lives, we do experience ourselves as human agents, at least some of the time. That is, we believe that our actions are in some sense, and for the most part, under our own jurisdiction, and that in these matters we largely decide for ourselves. Further, for the most part we experience ourselves, in Berlin's terms, as subjects in the face of multiple possible constraints. Berlin explains as follows:

I am the possessor of reason and will; I conceive of ends and I desire to pursue them; but if I am prevented from attaining them I no longer feel master of the situation. I may be prevented by the laws of nature, or by accidents, or the activities of men, or the effect, often undesigned, of human institutions.¹⁴

While we by and large experience ourselves as agents, we are also well aware of those times when our agency has been undermined. Our agency may be

¹³ I. Berlin, 'Two concepts of Liberty', in Berlin, I., *Four Essays on Liberty*, Oxford, Oxford University Press, 1969, p. 131.

¹⁴ *Ibid*, p. 135.

diminished by other agents as in cases of coercion or deception. We typically think that people who are brainwashed or threatened are unable to act intentionally or freely, and thus that their agency is in some way diminished. Our agency is also typically thought to be diminished by social or environmental constraints. In cases of repressive economic, social, political or cultural environments, for example, we think that people's ability to act freely or intentionally is restricted, and in extreme circumstances altogether undermined. Perhaps the most obvious cases in which we would think of agency as fundamentally undermined – where control over oneself and one's life are lost – are those of enslavement or imprisonment of the sort in which the human subject is treated deliberately as an object.

Viktor Frankl's famous *Man's Search for Meaning* is an existential exploration of the psychological condition of prisoners in concentration camps during the Holocaust. In that book, Frankl presents us with a popular, although extreme view, which I go on to criticise in this thesis, about the effect of such circumstances on our agency. According to Frankl, being imprisoned in a concentration camp constituted one of the most extreme situations in which human agency could be seen as fundamentally undermined. However, in Frankl's view, human agency is most fundamentally concerned with a kind of inner freedom or self-control, which he thinks remains available to us in even the most restrictive and oppressive circumstances, such as a concentration camp. At the very least, according to him, we have control over our internal mental and psychological states, and this is the kind of control with which external circumstances cannot interfere. He writes:

Does man have no choice of action in the face of such circumstances?

We can answer these questions from experience as well as on principle. The experiences of camp life show that man does have a choice of action. There were enough examples, often of a heroic nature, which proved that apathy could be overcome, irritability suppressed...

And there were always choices to make. Every day, every hour, offered the opportunity to make a decision, a decision which determined whether you would or would not submit to those powers which threatened to rob you of your very self, your *inner freedom*; which determined whether or not you would become the plaything of

circumstance, renouncing freedom and dignity to become moulded into the form of the typical inmate.¹⁵

This internal control focused picture of human agency has received attention in the analytic tradition of philosophy, and more in an important philosophically focussed biological discussion of agency.¹⁶ The picture of agency generated in this literature is explained in terms of self-control or behaviour modification – we exercise our agency when we exercise self-control or behaviour modification through the processes of critical assessment and evaluative deliberation. The most influential accounts of agency found in the philosophical tradition appear in the seminal works of Harry Frankfurt and Gary Watson. These philosophers have most influentially promoted views in which the attainment of self-control is constitutive of human agency, and the loss of self-control amounts to a breakdown, or undermining, of our agency.

In this chapter, I outline both a biological picture of self-control and the structural accounts of human agency as developed by Frankfurt and Watson. While Frankfurt and Watson's seminal self-control accounts provide us with a good foundation for an account of human agency, I will argue that they require some substantial revision and supplementation. In this chapter I put forward supplementations to the self-control view of human agency, drawing on work on the nature of human agency by philosophers such as Alfred Mele, Susan Wolf, Gerald Dworkin, and Mary Midgley as well as important psychological insights made by Sigmund Freud.

Following Frankfurt, Watson and Mele, in this chapter I will defend a self-control view of human agency, where self-control is construed as the ability to exercise sufficient motivational strength in order to ensure that intentional actions are performed in the face of competing desires. In addition, following Mele, I will argue that there are non-orthodox acts of self-control in which intentional action does not coincide with an agent's so called 'better judgements'.

¹⁵ V. Frankl, *Man's Search for Meaning, Ilse Lasche, I. trans.*, Boston, Massachusetts, Beacon Press, 2006, p. 74, my emphasis.

¹⁶ See N. Murphy and W. Brown, *Did My Neurons Make Me Do It?*. H. Frankfurt 'Freedom of the Will and the Concept of a Person'. H. Frankfurt, Identification and wholeheartedness'. H. Frankfurt, The Faintest Passion'. H. Frankfurt, 'Taking Ourselves Seriously'. G. Watson, 'Free Agency'. G. Watson, *Agency and Answerability*. A. Mele, *Irrationality*. G. Watson, *Autonomous Agents*. G. Watson. *Motivation and Agency*. G. Watson. *Backsliding*. G. Watson. 'Autonomy and Neuroscience'.

Nevertheless, I will argue that such actions are not to be seen as agency undermining. Following Wolf and Dworkin, I will argue that an agent's better judgements or intentional actions cannot be solely determined by third parties, such as in cases of mental manipulation or coercion. Further, I will argue that we must see human agency as something which emerges in varying degrees, and that being a human agent does not require that we exercise self-control at all times or in all aspects of our lives. Finally, following Midgley and Freud, I will argue that it is a necessary part of the self-control view of human agency that human agents experience internal conflict between their motivations and intentions.

1.1 Self-Direction, Self-Control and Human Agency

In this section I outline a plausible naturalist account of human agency which has found support in the biological sciences, drawing specifically on recent work by Nancey Murphy and Warren Brown. Though I will not defend naturalism itself in what follows in this thesis, I think it is important that an account of human agency can be formulated in broadly naturalistic terms. That is, we should be able to provide an account of human agency which does not require human agents to have supernatural or non-natural abilities. This, I think, is important if human agency is to be understood as distinct from any degree of agency we would attribute to non-human animals or supernatural beings such as celestial angels or god. Following this, I offer a view of human agency in what follows which takes biological as well as philosophical accounts seriously, and thus offer a robust view of human agency which naturalists could adopt or defend.

The kind of self-direction attributed to human agents on the naturalist's account accords in many ways with recent philosophical accounts of human agency, in which the distinctly human capacity Murphy and Brown describe can be understood in terms of self-control. I suggest here that we have reason to adopt something like the self-control view of human agency put forward by analytic philosophers such as Frankfurt and Watson, precisely because such views accord with the naturalistic picture of human agency presented by Murphy and Brown. On both Murphy and Brown's naturalist/biological view as well as

Frankfurt and Watson's self-control views, the ability for self-direction in order to achieve an intentional aim, rather than goal, is the defining characteristic of human agency. On the self-control view, the claim that a unique form of self-direction is constitutive of human agency is still explained by an appeal to teleology, as with all self-directed behaviour. Mele explains that "teleological explanations of human actions feature aims, goals, or purposes of human agents."¹⁷ For him, on a teleological view it is "a conceptual truth that an agent's acting intentionally at a time requires his having some relevant purpose or objective at that time."¹⁸ Similarly, on the biological picture as proposed by Murphy and Brown, all self-direction is necessarily goal-directed, and in light of this, self-direction is exhibited in all living organisms on a spectrum which ranges from virtually non-existent to significantly less sophisticated than the kind of self-direction human beings exhibit. Moreover, on both the philosophical and biological accounts, the high level of self-direction human beings exhibit is expressed in our ability to govern and control our actions and behaviour, and on such accounts this is what makes us actors or human agents as distinct from other organisms. While the self-direction required for human agency and the kind of self-direction exhibited by other organisms along the spectrum proposed by Murphy and Brown, the self-direction required for human agency is distinct from other types of agency in virtue of its biological complexity (seen in our brain and neural functioning) as well as its psychological complexity (seen in our ability to set systematic and reflective goals and values for ourselves).

Murphy and Brown have recently, and convincingly, combined cognitive science with philosophical thoughts on agency. They argue that on a continuum of the exhibited goal directed behaviour common to all organisms, human agents exhibit the most complex known example. For Murphy and Brown, the kind of activity exhibited by human agents is fundamentally rooted in the basic interaction between any organism and its environment (and this is true even for rudimentary organisms).¹⁹ Their project sketches what they call a "spectrum of

¹⁷ A. Mele, *Motivation and Agency*, p. 38.

¹⁸ *Ibid*, p. 164.

¹⁹ N. Murphy and W. Brown, *Did My Neurons Make Me Do It?*, p. 106.

forms of action,"²⁰ which reveals a gradient in the level of self-direction that we should attribute to organisms by taking into account the differences in the following: information about the environment which is available to the organism; the flexibility of behavioural responses which that organism is able to make to their environmental influences; and the complexity of the organism's evaluative systems. Murphy and Brown argue that human activity is positioned at the most complex extreme of this gradient, saying that:

conscious, deliberate, human action is a form of goal-directed activity involving the most complex analysis of environmental conditions, the largest number of response options, and the widest spatio-temporal scope for the evaluation of action possibilities and their consequences.²¹

First, they argue that while all activity is goal-directed – “aimed toward solutions to problems [goals] created by interactions between internal states and environmental circumstances,”²² what varies between organisms is the “complexity of environmental analysis,” which the organism can make, and the “flexibility of behavioural responses available to organisms.”²³

Second, Murphy and Brown argue that all biological activity has an evaluational component, which they call “action under evaluation.”²⁴ In most basic organisms, this simply refers to the fact that activity is part of a system in which behaviour can be corrected and modulated in relation to feedback received about the relationship between the outcomes of the behaviour and the goal of the activity. On Murphy and Brown's picture, this capacity is represented by a comparator that “determines if the goal is being reached and makes appropriate adjustments in the organisation of ongoing behaviour.”²⁵ What changes along the gradient of self-directed behaviour is the cause of the goals which direct the behaviour. For example, in a simple feedback system such as a thermostat, the goal is wholly set by something outside of the system itself. In contrast, the goals in simple biological organisms are set by evolutionary

²⁰ *Ibid*, p. 106.

²¹ *Ibid*, p. 107.

²² *Ibid*, p. 106.

²³ *Ibid*, p. 106.

²⁴ *Ibid*, p. 107.

²⁵ *Ibid*, p. 107.

processes. In more complex organisms, with comparatively more complex nervous systems, the behaviour-environment interactions become relatively more complex. This complexity is realised in “nested hierarchies of feedback and evaluation,”²⁶ and increases the ability of the organism to modify its own goals and moderate its own behaviour (through self-evaluation) which is initially set by biological determinism and the environment.²⁷

Third, Murphy and Brown argue that behaviours can be classified into three levels, which they claim to represent “milestones along a continuum of behavioural complexity”: namely, reflexive adaptability, unreflective adaptability and reflective adaptability. Reflexive behaviours include “responses that interact with environmental feedback, but according to fixed forms”.²⁸ For example, simple behaviour in single celled organisms, such as protozoa, with a singular form of behaviour have limited sensory mechanisms and limited possible responses to the environment, but still exhibit goal direction (“finding nutrients and avoiding toxins”²⁹), as well as action under evaluation (“modify the direction of swimming based on environmental feedback”³⁰). Unreflective adaptability refers to “behaviours that are modified in form by remembered environmental feedback.”³¹ Such behaviour would be exhibited in animals which are considered to ‘imitate’ other animals behaviour, or animals who are successfully able to engage in ‘trial and error’ tasks.³² Finally, reflective adaptability refers to “behaviours that are modified off-line by reflection using images or symbols.”³³ The use of images and symbols, as they are related to one another in a nexus of meaning is perhaps most widely associated with human behaviours, but there are less complex examples of reflective adaptability in the animal world. Murphy and Brown provide the following example:

For a monkey to know the meaning of an alerting call is for it to have a strong learned association between an iconic representation of the

²⁶ *Ibid*, p. 108.

²⁷ *Ibid*, p. 108.

²⁸ *Ibid*, p. 110.

²⁹ *Ibid*, p. 111.

³⁰ *Ibid*, p. 111.

³¹ *Ibid*, p. 110.

³² *Ibid*, p. 115.

³³ *Ibid*, p. 110.

particular call and an iconic representation of the threat of a leopard, for example.³⁴

While all three levels of behaviour are exhibited in both non-human and human organisms, the elaborate nested hierarchies of evaluative process, which are present in human cognition, can influence, modify or overrule behaviour arising at any of these levels – even those that are most basic. Murphy and Brown provide the following example:

coughing may be a simple reflex reaction to obstruction of the airways, but concert-goers have learned to restrain coughing until the end of the piece. So, when humans cough between pieces, there is an added level of conscious permission that complicates analysis of the behaviour.³⁵

Murphy and Brown's biological picture thus stresses the complexity of an important feature which is perhaps universally acknowledged as distinctive about an agent's mental life: the capacity for critical self-reflection and symbolic evaluation. From our own experience, it seems possible to say that a major part of what makes a human agent's mental life so complex is the fact that, as humans, we not only have primary or brute responses to the world, usually in the form of beliefs and desires, but that we are able to assess those basic responses, form opinions about them, and often even change them. Murphy and Brown claim that while this capacity, exemplified in what they term the comparator, is present in all organisms, the human comparator is the most complex example because it results in the most complex examples of behaviour. When critically reflecting on our basic responses to the world, we evaluate whether or not we are motivated by our basic responses by checking those responses against our values, commitments, projects, aims and goals. Such processes of critical self-reflection and symbolic evaluation result in the most complex form of reflective adaptability. What Murphy and Brown explain as the comparator is what we typically take ourselves to be when we are exercising our agency. That is, when we refer to our own experience of ourselves as agents, what we are typically

³⁴ *Ibid*, p. 121.

³⁵ *Ibid*, p. 110.

referring to is an experience of ourselves as comparators, i.e. active evaluators of our experiences.

Similarly, it is our evaluation of our experiences which is central in recent philosophical accounts of human agency. The crucial capacities for self-reflection and symbolic evaluation are discussed in this literature in terms of a capacity for self-control. It is control over the self born from critical self-reflection and symbolic evaluation in particular that is taken to be constitutive of autonomous or free agency (encapsulated by what I refer to in this thesis as human agency). If the biological account of human agency is sufficiently similar to the self-control view of human agency presented in the philosophical literature, it would certainly give us good reason to adopt the self-control view, particularly if we are committed to naturalism. On the naturalist view, self-control is a complex example of the kind of intentionality exhibited by all other living things along a spectrum. Human agency remains on this spectrum, and yet is clearly distinct from other types of agency along that spectrum in virtue of the capacity for self-control. Further, on the biological view, human agency remains distinct from the kind of agency perhaps exhibited by supernatural and disembodied beings, who may arguably hold a place on the self-direction spectrum, but who are to be considered distinct from human agents since human agents are necessarily embodied. In what follows I outline two influential accounts of human agency from the analytic tradition which centre on self-control and which are compatible with the biological view in the above respects.

Two plausible and important philosophical accounts of human agency in which self-control plays a central role have been proposed by Harry Frankfurt and Gary Watson. Central to these accounts is the idea that to be in control of our own motivations, and the actions we take as a result, is what marks us out as distinct from the rest of the animal kingdom and is what makes us human agents. Both Frankfurt and Watson's accounts highlight the role of self-control in human agency. And to this end, they outline the bare minimum necessary structural features of the mind, which allow us to control our initial responses to the world, as well as allowing us to perform actions that are calculated and controlled, rather than being largely out of our control. As their starting points, both Frankfurt and Watson take the basic assumption that as human agents we

characteristically not only have certain overarching or guiding principles and values (often referred to as 'our better judgement'), but that we are able to assert some form of control over our choices and actions when our better judgement effectively guides our actions. On their accounts, when our better judgement guides our actions in spite of any motivation we may have to act to the contrary, we exercise the kind of self-control necessary for human agency.³⁶

In his paper 'Freedom of the Will and the Concept of a Person' Frankfurt distinguishes between what he calls first and second order desires; and further between first or second order desires and what he calls second order volitions. He claims that only once our second order volitions are aligned with our first order desires, both of which constitute our will, can we be seen to exercise the kind of self-control that is constitutive of human agency.

On Frankfurt's account, first order desires are largely determined by our initial response to the world. Typically, an agent will have a first order desire in the form of basic want as expressed in, for example, "I want to go for a jog now." This kind of first order desire is often experienced in competition with other first order desires. For instance, an agent desiring to go out for a jog now might also find herself with the competing desire to watch a documentary now. In the face of this internal competition, some desires rather than others will successfully move us to perform corresponding actions at any given time. According to Frankfurt, first order desires that are effective in this way, or that would be effective if we were not physically prevented from acting in this way, are taken to constitute our will.³⁷ For Frankfurt, a person's will is also comprised of those desires "by which he will or would be motivated when or if he acts."³⁸

Second order desires are the desires we have as a product of critical reflection and evaluative assessment of our first order desires, and thus have as their content a reflective assessment of a first order desire. For example, when

³⁶ In important clarification of terminology is important here. Frankfurt refers to what I am here describing as human agency in terms of freedom of the will, while Watson refers to it in terms of free agency. The account of human agency I am putting forward based on the work of Frankfurt and Watson should be seen as neutral with regard to what position one may take in the free will debate.

³⁷ H. Frankfurt, 'Freedom of the Will and the Concept of a Person', p. 8.

³⁸ *Ibid*, p. 8.

critically reflecting on our first order desire to go for a jog now, we may reflectively evaluate this desire and arrive at a higher order desire such as “*I want to want to go for a jog now*” or “*I want to want to watch a documentary now*”. Second order desires are, according to Frankfurt, the product of that characteristically human capacity for critical self-reflection on our initial responses to the world – or our first order desires – and are meant to explain the ability of human agents to have intentions and preferences about their own will. That is, by wanting our will to be different to what it happens to be, we are able to some extent to regulate it.

Moreover, some second order desires are to be seen as what Frankfurt calls second order volitions. According to Frankfurt, a second order desire can be seen as a second order volition when, with regards to a specific first order desire, we want that first order desire to be our will. A second order volition, then, is a desire of the second order which corresponds with a desire of the first order that effectively moves us to action.³⁹ For Frankfurt, second order volitions are agency conferring – that is, when our actual will mirrors the will which is endorsed by our higher order volition, we can be said to have the will *we want to have*. Frankfurt maintains that by having the will we want to have in this sense, our will is free in the way required for human agency because it is the result of our successful efforts at self-control. Frankfurt claims the following:

The statement that a person enjoys freedom of the will means [...] that he is free to will what he wants to will, or to have the will he wants [...]. It is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will.⁴⁰

Frankfurt further notes that when our will mirrors our second order volitions, we are free from any factors that could be seen as agency undermining precisely because we are moved to action by something that we have endorsed for ourselves. Frankfurt says:

³⁹ *Ibid*, p. 10.

⁴⁰ *Ibid*, p. 15.

... since the will that moved him to action was his will because he wanted it to be, he cannot claim that his will was forced upon him or that he was a passive bystander to its constitution.⁴¹

So, for Frankfurt, it is possible to say that we have *control* over our actions because we are human agents in the sense that we are able to exercise *self-control*. Self-control is achieved, on Frankfurt's account, when our second order volitions mirror our will. And, for him, we can be said to have the will that *we want* to have precisely because it is endorsed by us. According to Frankfurt's initial account, then, what it means to have a free will is that our will is under control, in the sense that it is endorsed by our higher-order mental elements, which themselves have been formed in the light of rational critical reflection on ourselves.

On Watson's view, which is similar but not identical to Frankfurt's, what is distinctive about human subjects is their ability to make rational evaluations about themselves and their situation in the world, and to form values in light of these deliberations which guide their actions when they are free agents.⁴² On Watson's account, the values we have, as human agents, are a reflection of who we are, or are constitutive of what we could call our true or real selves. Thus, for Watson, when our actions reflect our values, they are attributable to us in the deep sense of attributability which is thought to underlie freedom, or free agency, precisely because our actions have been up to us in the relevant sense. Most importantly, for Watson, when we are able to act in accordance with our values in the face of various obstacles or constraints (such as those desires which, if acted upon, would undermine one or more of our rationally chosen values), we exercise that capacity which is the cornerstone of human agency, namely self-control. Similarly to Frankfurt's account, then, Watson's account tries to capture an intuition about human agency that focuses on self-control. This intuition is expressed succinctly by Richard Moran when he explains the following:

⁴¹ *Ibid*, p. 18.

⁴² G. Watson, 'Free Agency'.

It is the normal expectation of the person, as well as a rational demand made upon him, that the question of what he actually does desire should be dependent [...] on his assessment of the desire and grounds he has for it. For the person himself, then, his motivated desire is not a brute empirical phenomenon he must simply accommodate, like some other facet of reality he confronts. For this sort of desire, as a “judgment-sensitive” attitude, owes its existence (as an empirical psychological fact) to his own deliberations and overall assessment of his situation.⁴³

Models of human agency, such as Watson’s, are sometimes referred to as partitioning accounts, which explain our agency as self-control via the notion of a partition within the psyche between motivational systems (consisting of the considerations which prompt a person to action) and evaluational systems (consisting of overarching principles and value systems). Watson claims that if it is possible for a discrepancy to exist between these two independent sources of motivation (desire and value), that it is possible for an action to be either free or unfree. Watson explains:

The problem of free action arises because what one desires may not be what one values, and what one most values may not be what one is finally moved to get.⁴⁴

The question of whether or not we act freely is thus, in Watson’s opinion, focused on how we might be in control over what motivates us *effectively* – that is to say, what moves us all the way to action. Since there are, according to Watson, two sources or springs of motivation (namely desires and values), there arises the possibility of our being motivated to follow two distinct and potentially contradictory courses of action simultaneously – both from within our own mental complex. If either one of these distinct or contradictory motivations were to be effective in action, one motivation would be trumped by another motivation held by that same person. I may, for example, desire and be motivated to have chocolate cake, while at the same time value a healthy lifestyle and be motivated not to eat refined sugars. According to Watson, we act as free agents, or human agents, only when our values and principles (those motivations

⁴³ R. Moran, *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton, Princeton University Press, 2000, p. 115.

⁴⁴ G. Watson, ‘Free Agency’.

arising from our valuation system) are effective motivations, which is to say when they move us all the way to action. Our actions, then, are free when our motivational system mirrors our evaluational system. Watson explains, as follows:

The possibility of unfree action consists in the fact that an agent's evaluational system and motivational system may not completely coincide. Those systems harmonize to the extent that what determines the agent's all things considered judgments also determines his actions [...]. The free agent has the capacity to translate his values into action; his actions flow from his evaluational system.⁴⁵

The suggestion of such a partition allows Watson to draw a distinction between free and unfree action. As such, this account helps us to establish precisely why we can distinguish between individuals who are coerced or compelled, and those who are suitably free from constraints which would otherwise diminish their agency. In cases in which a person has been coerced, or in some way compelled, his actions are unfree precisely because what he is motivated to do is in tension with what he has, himself, rationally evaluated to be the best course of action. Watson explains the way in which a kleptomaniac is unfree in the following way:

What is distinctive about such compulsive behavior, I would argue, is that the desires and emotions in question are more or less radically independent of the evaluational systems of these agents.⁴⁶

What Watson makes explicit in the conditions he provides for free agency is that when we perform an action freely, it means that we are moved to action by our evaluating rational selves, or by our evaluational system. That is, the values and principles that form part of our evaluational system are a product of our own rational evaluations. Human agency, then, on Watson's account, requires us to exercise control over our actions by always acting on our values or principles, or to exercise self-control. According to Watson, human agency is constituted by three factors. First, human agency requires that we have a point of view from which to judge the world, which can be said to constitute *our* point of view. Second, human agency requires that our point of view is to be identified with

⁴⁵ *Ibid*, pp. 215-216.

⁴⁶ *Ibid*, p. 220.

what he calls our evaluational system, as Watson explains: “One’s evaluational system may be said to constitute one’s standpoint, the point of view from which one judges the world.”⁴⁷ Finally, human agency requires that we, in fact, do act on the dictates of our evaluational system.

1.2 Constraint, Sanity and Procedural Independence

In this section I argue that the classic accounts of the self-control view of human agency, provided by Frankfurt and Watson, require clarification. Apart from human agents being required to reflectively assess and endorse certain of their motivations over others, I will argue in this section that a human agent’s better judgements or intentional actions cannot be solely determined by third parties. Someone whose better judgement or intentional actions are determined solely by third parties are, I suggest, for the most part to be seen as lacking the kind of critical or evaluative freedom necessary to exercise human agency. Further, I suggest that being a human agent does not require that we exercise self-control at all times, or in all aspects of our lives.

Self-control consists in guiding our actions and behaviours in light of our better judgement (or our critical reflections and evaluations), but there are times when certain constraints may trump our ability to form better judgements or to make our critical assessments in ways which would not allow us to exercise the degree of self-control necessary for human agency. All of us are subject to influence from external forces when assessing and choosing from a number of possible courses of action we might take, and so to some extent we must recognise that the self-control necessary for human agency can never be wholly free from all external influence and constraint. For example, we are concerned with how other people will view our choices and actions, and this can have some effect on what we decide to do. But, at least some of the time, the external constraints we face play a largely insidious role – they undermine our ability to form intentions and choose courses of action which are in any relevant sense our own. In situations where we are so constrained that we act on motivations which are not in the

⁴⁷ *Ibid*, p. 216.

relevant sense our own (that is, as the product of our own evaluations and assessments), our freedom is typically taken to have been undermined and we are no longer responsible for ourselves in the sense required for human agency. We do not typically consider people who are subject to mental manipulation or coercion to be free or accountable for their actions because they have been unable to either act on or form their own better judgments or critical evaluations. We typically think that people who are brainwashed, or whose lives or loved ones are threatened, have had their agency most basically undermined; and, in such cases, people are no longer seen as controlling their own actions in any sense relevant for accountability. Rather, their actions are seen as controlled by their brainwasher or coercer.

While to have self-control is to be in control of oneself, there is certainly more to being in control of oneself than having and exhibiting the power to master motivation that is contrary to one's better judgement. If our better judgements rest on values generated and maintained by brainwashing, or are influenced by nefarious ideologies that we would not endorse if only we could see them clearly, we would exercise self-control according to Frankfurt and Watson's views, but we do not seem to be in control of ourselves in a broader sense. In a broader sense such a person would be ruled, as Mele argues, ultimately, not by his 'self' but rather by his brainwasher, the insidious effect of ideology, or society at large.⁴⁸ To be in the relevant sense accountable for our actions as agents, we must exercise self-control in this broader sense. However, the broader sense of self-control in which accountability is at stake should not be thought of here as also providing a basis for the ascription of moral responsibility.

In her paper "Sanity and the Metaphysics of Responsibility,"⁴⁹ Susan Wolf offers the example of JoJo, "the favourite son of Jo the First, an evil and sadistic dictator of a small, undeveloped country."⁵⁰ In Wolf's example, JoJo spends a great deal of time in the company of his father, who shows his son a sense of favouritism. JoJo thus "takes his father as a role model and develops values very

⁴⁸ A. Mele, *Irrationality*, p. 61.

⁴⁹ S. Wolf, 'Sanity and the Metaphysics of Responsibility', in Schoeman, F. (ed.), *Responsibility, Character and the Emotions: New Essays in Moral Psychology*, Cambridge University Press, 1998.

⁵⁰ *Ibid*, p. 367.

much like his Dad's."⁵¹ According to Wolf, when JoJo acts as his father did, for example by "sending people to prison or to death or to torture chambers on the basis of whim," he is not constrained or coerced into action because his desires are wholly his own in Frankfurt and Watson's sense – that is, his "actions are controlled by his desires and [his] desires are the desires he wants to have."⁵² However, and importantly, Wolf points out that given JoJo's circumstances he cannot truly be said to be, in the relevant sense, accountable for his values that inform his better judgement. Wolf explains the following:

In light of JoJo's heritage and upbringing – both of which he was powerless to control – it is dubious at best that he should be regarded as responsible for what he does. It is unclear whether anyone with a childhood such as his could have developed into anything but the twisted and perverse sort of person that he has become.⁵³

Following Wolf's view, the self-control view of human agency, as established by Frankfurt and Watson requires an additional element which would allow us to distinguish between those persons who are able to freely form their better judgement (or exercise critical reflection), and those who, although able to act in accordance with their better judgement, are unable to form their better judgement (or exercise critical reflection) freely.

Here, I agree with Wolf that Frankfurt's and Watson's self-control views of human agency require supplementation. Given the complexity of the effects of constraints on our mental lives, an account of human agency must allow us to distinguish between JoJo and someone who can be seen as accountable for their action in the relevant sense. A person who is seen as accountable for their actions in the relevant sense here is someone who is judged by broader society to have acted without any agency undermining constraints influencing the chosen course of action. For someone to be held accountable, they need not be held morally responsible, although we may award praise or prescribe certain rehabilitative measures to someone who we would consider to be accountable. Any supplementation to Frankfurt and Watson's self-control view of human agency, then, should explain how JoJo differs from a person who we would hold

⁵¹ *Ibid.*

⁵² *Ibid.*, p. 368.

⁵³ *Ibid.*

accountable, and it should explain why we would attribute accountability to such a person rather than JoJo.

A convincing candidate for just such supplementation has been proposed by Gerald Dworkin. In his paper 'Autonomy and Behaviour Control',⁵⁴ Dworkin introduces the condition of 'procedural independence' as fundamentally important to discussions about the nature of human agency, and he continues to argue for the relevance of 'procedural independence' in his book *The Theory and Practice of Autonomy*.⁵⁵ According to Dworkin, as individuals, we lack 'procedural independence' if we are unable to view our own situation clearly, for example, if we are in a threatening situation. In such cases, Dworkin points out that an "individual may identify or approve of his motivational structure because of an inability to view in a critical and rational manner his situation."⁵⁶ Dworkin also argues that we lack 'procedural independence' if we are subject to external forces which wholly determine our better judgement. He explains:

the identification with his motivations, or the choice of the type of person he wants to be, may have been produced by manipulation, deception, withholding of relevant information, and so on. It may have been influenced in decisive ways by others in such a fashion that we are not prepared to think of it as being his own choice. I shall call this a lack of procedural independence.⁵⁷

If we are to be considered accountable for our actions in the sense relevant for human agency, then, in addition to the internal conditions involving what motivates us to act from within our own mental complex, our own better judgement and our critical reflection must also be suitably free from external influences which might be considered to undermine agency. Drawing on Dworkin's negative account, what is needed to bolster Frankfurt and Watson's accounts is the following twofold 'procedural independence' condition in which a human agent's evaluative system is, firstly, not undermined by a faulty rational or critical ability, and secondly, is not formed purely as stipulated by some third

⁵⁴ G. Dworkin, 'Autonomy and Behaviour Control', *Hastings Center Report*, vol. 6, 1976, pp. 23-28.

⁵⁵ G. Dworkin, *The Theory and Practice of Autonomy*, Cambridge, Cambridge University Press, 1988.

⁵⁶ G. Dworkin, 'Autonomy and Behaviour Control', p. 25.

⁵⁷ *Ibid*, p. 25.

party, as in the cases of deception and manipulation, as well as in cases where certain relevant information is withheld.

Wolf, however, proposes an alternative plausible supplementation for the self-control view of human agency in the form of an epistemic requirement for sanity. While Dworkin argues that what we need for accountability is another form of control over and above the kind of self-control described by Watson (in the form of procedural independence), Wolf denies that what we need for accountability is more control. Wolf argues that we do not need more or stronger control over ourselves, rather what is needed in addition to self-control is an epistemic condition in the form of sanity, which is utterly unlike a form of control. Over and above the ability to exercise self-control, Wolf proposes that human agents must also meet a sanity condition, which acknowledges that there is a particular kind of connection between our selves and the world. JoJo's problem is that his upbringing is so bad that he has significantly lost the ability to grasp what is the case and what is not. His connection to the world has crucially been severed, and he is no longer sane (in Wolf's sense). JoJo can explicitly endorse his father's views, but he does so because he is fundamentally insane, which is to say that his father's worldview is engraved into JoJo's psyche. Wolf explains as follows:

in order to be responsible an agent must be *sane*. It is not ordinarily in our power to determine whether or not we are sane. Most of us, it would seem, are lucky, but some are not. Moreover, being sane does not necessarily mean that one has any type of power or control an insane person lacks. Some insane people, like JoJo and some actual political leaders who resemble him, may have even complete control over their actions, and even complete control over their acting selves. The desire to be sane is thus not a desire for another form of control; it is rather a desire that one's self be connected to the world in a certain way – we could even say it is a desire that one's self be *controlled by* the world in certain ways and not in others.⁵⁸

For Wolf then, it is not enough for us to say that human agency requires the kind of motivational self-control proposed by Frankfurt and Watson, human agency requires that we are suitably free from constraints that prevent us from properly assessing the goals and values which guide and inform our better judgments. It is our agency itself which is undermined, for Wolf, when our situation in the world

⁵⁸ S. Wolf, 'Sanity and the Metaphysics of Responsibility', p. 368.

prevents us from being sane, that is, prevents us from having an ‘accurate conception of the world’ and from having ‘blind or distorted forms of responses’ to the world. Wolf explains further as follows:

Insofar as one’s desire to be sane involves a desire to know what one is doing – or more generally, a desire to live in the real world – it is a desire to be a controlled [...] by perceptions and sound reasoning that produce an accurate conception of the world, rather than blind or distorted forms of response [...]. [Furthermore,] one’s hope is that one’s values be controlled by processes that afford accurate conceptions of the world.⁵⁹

Sanity, then, is explicitly not a matter of control for Wolf. It’s a matter of seeing the world correctly. On Wolf’s account the difference between JoJo and the rest of us is not a matter of us having procedural independence and JoJo not having it. The difference is that JoJo is unable to see the world clearly, or to understand his proper connection to the world. It’s explicitly an epistemic problem, not a problem of control. Importantly, on Wolf’s account it is the content of what JoJo is exposed to that is the problem, not the extent or manner of the exposure. One attractive feature of Wolf’s view is that it has the consequence that ‘normal’ people who simply inherit their world view from their parents or communities (like many of us) come out as sane agents.

Wolf’s sanity condition, then, certainly provides a plausible candidate for supplementation to the self-control view established by Frankfurt and Watson, but it is important to note that through the sanity condition Wolf also aims at an understanding of human agency as a substantive rather than a purely formal concept. Like Frankfurt and Watson, I favour a formal concept of human agency – seeing Normativity associated with personhood, as I will argue in the second part of this thesis. Further, the deeper concern about the kind of control we have over our ability to critically reflect on our values and form our better judgment, can, I think, be accommodated in the formal self-control view of human agency without requiring us to understand human agency as a substantive notion if we follow Dworkin’s suggestion of procedural independence. My favouring of a formal over a substantive account of human agency, and thus my preference for Dworkin’s account, does not of course settle the dispute between Wolf and Dworkin on

⁵⁹ *Ibid*, pp. 368-369.

whether procedural independence or sanity is the right sort of further ingredient needed for an account of human agency in addition to Frankfurtian/Watsonian self-control. The central concerns about human agency and personhood which I go on to address in this thesis are not influenced by whether we adopt Wolf's or Dworkin's approach to broadening our understanding of when an agent can be said to exercise self-control in the sense relevant for accountability – either supplementation could satisfactorily do the trick. Importantly, for my purposes here, on both Dworkin's and Wolf's account, a human agent whose better judgement or intentional actions are determined solely by third parties are, for the most part, to be seen as lacking the kind of critical or evaluative freedom necessary to exercise human agency, or to be accountable for their actions.

Both Wolf's and Dworkin's additions to the classic self-control view of human agency given by Frankfurt and Watson, however, are more complicated than they might at first have seemed. It is far more difficult for us to delineate between cases in which accountability is lost, or in which 'procedural independence' or 'sanity' has been undermined, from those in which it is not.

In everyday life, accountability is often formally assigned by a lawyer or jury in a court of law, and a country's laws are to a large extent a reflection of the kinds of constraint which that society takes to undermine accountability.

Similarly, popular literature and film often portray cases in which the audience or reader is asked to consider whether or not the protagonist is subjected to constraints which we would take to undermine their accountability. We typically tend not to hold accountable characters who have been coerced into action or who face science-fiction style scenarios in which they have been subject to extreme mental manipulation. For example, in John Frankenheimer's thriller *The Manchurian Candidate*, we sympathise with the protagonist Raymond Shaw, who is brainwashed and subject to mental manipulation such that we take him to have been 'programmed' as a Soviet assassin in order to target a presidential candidate, rather than as someone who is accountable for having devised, of their own accord, a devious plan to commit murder.⁶⁰

⁶⁰ *The Manchurian Candidate*, dir. John Frankenheimer, USA, M. C. Productions, 1962.

Further, we generally take people living under oppressive regimes⁶¹ to have a diminished sense of accountability for their choices and actions. History teaches us through illustrative cases – consider the Nuremburg Trials and the South African Truth and Reconciliation Commission (TRC) – that we tend to respond to those who commit various crimes in extreme social conditions in similar ways to those who have been coerced or manipulated – at least some of the time. In the Nuremburg trials, those in positions of power were punished, not the deluded German youth. Similarly, the hearings of the TRC served to promote the kind of understanding of the constraints faced by many of the perpetrators of crimes against humanity under the Apartheid regime, and through this understanding, to favour merciful mitigation in the form of amnesty over vengeance or strict retributivism. The TRC, moreover, sought to transform the actual structural and institutional problems which allowed for the perpetuation of human rights violations rather than simply dealing with individual perpetrators. In support of this intuitively attractive view of the limits of culpability, Simone de Beauvoir famously claimed that we do not:

hate those youthful sixteen-year-old followers of Hitler in whom Nazism affirmed itself with such violence, but who never had the possibility of criticizing it. One reeducates children, the ignorant, those populations that are ill informed; one does not punish them.⁶²

On Wolf's account, those very real cases of the deluded German or South African youth are analogous to her fictional case of JoJo. She claims:

we give less than full responsibility to persons who, though acting badly, act in ways that are strongly encouraged by their societies – the slave owners of the 1850s, the Nazis of the 1930s, and many male chauvinists of our fathers' generation [...]. If we think that the agents could not help but be mistaken about their values, we do not blame them for the actions those values inspired.⁶³

⁶¹ This is true for both the 'oppressed' and the 'oppressors'. For example, under the Apartheid regime, both black and white South Africans were subject to external constraints which by and large undermined their ability to formulate for themselves their 'better judgment' in the absence of severe distorting constraints.

⁶² S. Beauvoir, 'An eye for an Eye', in Simons, M. (ed.), *Simone de Beauvoir: Philosophical Writings*, Urbana, University of Illinois Press, 2004, p. 257.

⁶³ S. Wolf, 'Sanity and the Metaphysics of Responsibility', p. 369.

Further, we also tend to think of certain environmental constraints as agency undermining. We take people who find themselves in life threatening, do-or-die situations to be less responsible for their choices and actions. We are, for example, less inclined assign accountability to people who engage in so-called 'survival cannibalism' because we view their choices and actions as radically constrained – we see them as unable to exercise the kind of freedom associated with agency. Anthropologically speaking, the prevailing view in western society is that those engaged in 'survival cannibalism' are exempt from accountability for their choices and actions. In response to famous cases in which survivors of plane accidents in remote areas are forced to eat the bodies of their dead fellow passengers, such as the situation in 1972 in which 16 air accident survivors were stranded in the Andes, were deemed exempt by the public from accountability.

It is important to point out, however, that there is by no means broad consensus in any given society, let alone across different societies, about what the limits of accountability are. The effects of some external constraints on our agency are less intuitively easy for us to explain: where we draw the line when establishing whether certain factors preclude our holding someone accountable is largely blurred and sometimes flexible. Within the legal world there is much debate and argument amongst lawyers and in courtrooms about whether accountability should be assigned to a defendant, given the particulars of their situation. Similarly, popular literature and some writings in moral philosophy have challenged us to question whether all actions conducted in situations of circumstantial constraint are really to be forgiven. In the award-winning novel *Life of Pi*⁶⁴, we are asked to imagine a ship-wreck situation, quite dissimilar to the 1972 plane accident situation, in a salient way: in the *Life of Pi* situation, our protagonist is stranded on a life-boat with a character who many would deem a moral monster because, as the story could be interpreted, he murders fellow passengers in order to engage in 'survival cannibalism.'

In many cases certain choices made or actions undertaken by people in situations of extreme social constraint provoke a condemnatory reaction that is shared in western society. Such condemnatory response is largely a reaction to the kinds of actions in which these people have engaged. The perhaps

⁶⁴ Y. Martel, *Life of Pi*, New York, Harcourt Inc., 2001.

overwhelming social constraint present in Nazi Germany or Apartheid South Africa does little to conciliate the perceived monstrosity of characters such as Adolf Eichmann or Hendrik Verwoerd – and even to those lower ranking soldiers and civilians who committed seemingly unforgivable acts of gross human rights violations. Though we typically think that a person’s agency may be undermined by certain kinds of social constraints, it also appears that there are some firmly held intuitive boundaries to the range of choices and actions for which a person can be held accountable, even in the face of severely restrictive social constraints. Our intuitions about the extent of mitigation attained when someone is constrained by external circumstances vary, it would seem, not exclusively according to the level of constraint involved, but with respect to the kinds of choices and actions people make in these situations.

Importantly, what our divided views about the role that external circumstances play in either diminishing or affirming accountability should tell us is that, for the most part, we intuitively struggle to delineate any comprehensive boundaries for our own agency in terms of accountability. Perhaps, it might be argued here, we ought simply to give up on some of our intuitions and simply draw the decisive boundaries. But this bullet is hard to bite. While analytic philosophical pictures have significantly helped us to understand the nature of human agency, insights from this tradition must be supplemented with an account which takes the lived experience of agency seriously, as well as an account in which culpability is socially determined. In taking our own intuitions about the boundaries of our agency seriously, we can, I think, arrive at some important insights about the nature of agency which supplement the analytic picture – and I go on to discuss these insights further in the second chapter. Drawing on what I have discussed in this section one insight seems important to point out here: in our experience, agency is the kind of thing that exists in degrees. We can be sometimes more or less subject to forces which seem to diminish our accountability for our choices and actions in the eyes of society. Mele argues that “[s]elf-control may be either regional or global,” and that “it comes in degrees.”⁶⁵ I think that Mele rightly argues the following:

⁶⁵ A. Mele, *Backsliding*, p. 93.

The fact that a scholar exhibits remarkable self-control in adhering to the demanding work schedule that he judges best for himself does not preclude his being weak-willed about eating. He may be self-controlled in one “region” of his life and weak-willed in another. And some self-controlled individuals apparently are more self-controlled than others. Agents who have global [...] self-control in all regions of their lives [...] would be particularly remarkable, if, in every region, their self-control considerably exceeded that of most people.⁶⁶

It is a common experience as agents to find that while we may be self-controlled in some areas of our lives, we might completely lack self-control in others. And this should not be seen as a failure on our part. In fact, it might be important for us, within certain limitations, to relinquish self-control and allow ourselves to be guided by chance. Our participation in some of life’s most pleasurable activities, such as dancing or sex, are enhanced, rather than impeded, by ‘letting go’ and being out of control – guided in a mad frenzy by the beat of the music, or our lover’s touch. In the second part of this thesis I return to this idea, arguing that self-governance involves knowing when to let go of control.

1.3 A Disagreement about Our Better Judgement

In this section, following Mele, and contra Frankfurt and Watson, I argue that intentional actions need not always be in the service of our reflective better judgements. While Frankfurt and Watson maintain that the self-control required for human agency rests on our ability to act in accordance with our better judgements, I argue that an agent is more properly understood as someone who exercises self-control by mustering the motivational strength to perform intentional action in the face of competing motivations.

Contrary to the views of self-control put forward by Frankfurt and Watson, Mele has, I think rightly, argued that a view in which “all intentional actions contrary to the agent’s better judgment are unfree,”⁶⁷ is implausible. Mele summarises Watson’s argument as follows: “An agent’s succumbing to a desire contrary to his better judgment cannot be explained by his choosing not to

⁶⁶ *Ibid*, p. 93.

⁶⁷ *Ibid*, p.336.

resist, nor by his making a culpably insufficient effort to resist.”⁶⁸ On such a view, as Mele interprets it, there is “[o]nly one explanation [that] remains: the agent was *unable* to resist.”⁶⁹ In other words, cases in which we act against our better judgement are cases in which we have no self-control and are therefore not acting as agents. In contrast to Watson’s view, Mele argues, first, that in our experience we are able to – and do – perform actions against our own better judgement, and second, that in such cases we do not always think that we have been constrained in any way. That is, we still think that we have acted freely or exercised our agency. Mele asks the following question: “Why do ordinary folks believe that there are (in my terminology, not theirs) core weak-willed actions?”⁷⁰ And he provides the following answer:

Why do ordinary folks believe that there are (in my terminology, not theirs) core weak-willed actions? Presumably, largely because they take themselves to have first hand experience of such actions and partly because some of their observations of—and conversations with—others indicate to them that they are not alone in this. It seems to them that, occasionally, they perform, in the absence of compulsion, intentional actions that are contrary to their *J*. They have, at such times, no feeling of being compelled. Indeed, they may feel free and uncompelled, and they may believe at the time of action that they are able to do what they judge best at the time.⁷¹

Mele defends the popularly held view that a human agent can intentionally act against her own better judgement, and advances plausible cases in defence of this view. On Mele’s view, forming a better judgement on the basis of our critical reflection does not necessarily include a corresponding motivation or intention to act. For Mele, “it is very plausible that some agents who judge at *t* that it is best to *A* at *t* do not have a corresponding intention at *t* and intend a course of action that is at odds with that judgment.”⁷² He offers a first case, in brief, of the crack addict who “might judge (believe) it best not to use crack now, lack an intention not to use it now, intend to use it now, and intentionally use it now.”⁷³ Recall that on Frankfurt and Watson’s views, when we act on desires or motivations that

⁶⁸ *Ibid*, p. 336.

⁶⁹ *Ibid*, p. 336.

⁷⁰ *Ibid*, pp. 43.

⁷¹ *Ibid*, pp. 43-44.

⁷² *Ibid*, p. 40.

⁷³ *Ibid*, p. 40.

run counter to our better judgement, our actions are not free and we are not exercising the self-control required for human agency. In the case of the crack addict, Watson takes the addict to be 'unfree' because, according to him "it is plain that in some situations [like that of the crack addict] the motivational systems of human beings exhibit an independence from their values which is inconsistent with free agency."⁷⁴ Watson says of the compulsive, such as the kleptomaniac, that "it is because his desires express themselves independently of his evaluational judgments that we tend to think of his actions as unfree."⁷⁵ For Frankfurt, the case of the addict presents us with what he calls the "unwilling" or non "wanton" addict, who forms a second order volition not to take drugs. In such a case, Frankfurt argues that:

the unwilling addict may meaningfully make the analytically puzzling statements that the force moving him to take the drug is a force other than his own, and that it is not of his own free will, but rather against his will that this force moves him to take it.⁷⁶

On Mele's view, however, we can "still choose, decide or intend" to act against our own better judgement. He advances a plausible case in favour of this view by presenting that of Drew, the weak-willed drinker. He illustrates Drew's case as follows:

Drew, who has had one shot of whiskey and needs to drive home soon, judges it best to switch now to coffee but neither chooses, decides, nor intends accordingly. She intentionally takes another whiskey.⁷⁷

According to Mele, in the case of Drew it is easy to see how her drinking another shot of whiskey is both intentional and free, *and* simultaneously against her better judgement. He explains the case of Drew in further detail as follows:

She believes both that she can switch from whiskey to coffee now and that she can have another whiskey now instead; in Drew's opinion, it is up to her which of these she does. She knows that it is risky to drive under the influence of two shots and she judges that, in light of the risk, her reason

⁷⁴ G. Watson, *Agency and Answerability*, p. 31.

⁷⁵ *Ibid*, p. 31.

⁷⁶ H. Frankfurt, 'Freedom of the Will and the Concept of a Person', p. 13.

⁷⁷ A. Mele, *Backsliding*, p. 39.

for having a second shot—that she would enjoy it—does not justify having one. Although Drew judges that she should switch to coffee, she thinks “I’ve had a bit too much to drink before and all has gone well. I really should switch to coffee, but I’ll indulge myself. Just one more shot, then a cup of coffee, then I’ll drive home.” Still believing that it would be best to switch to coffee now, Drew decides to drink another shot and drinks one. She does not feel compelled to drink. She feels that she is deciding freely and that she is freely drinking the whiskey.⁷⁸

I agree with Mele that the case of Drew illustrates what is actually a commonly held view on which it is possible to act freely even when one does not act in accordance with one’s better judgement. Mele claims that his view of self-control best fits the common sense view of self-control. In a study of undergraduate students conducted by Mele himself at Florida State University, approximately half of the participants agreed that, in a case similar to that of Drew’s case (as well as the case of Bruce which I will discuss below) a person should be seen to have exercised self-control.⁷⁹ Despite the disagreement about the role of our better judgement in exercising the kind of self-control associated with agency, Mele still subscribes to a view in which self-control is central to human agency.

So why is the disagreement about better judgement so important? I think that this disagreement is important for understanding what a plausible self-control account of human agency would look like. On Mele’s view, a version of self-control is still central to understanding free action or human agency. Mele distinguishes here, however, between what he calls “orthodox” cases of self-control, in which “exercises of self-control serve the agent’s better judgment,”⁸⁰ and “unorthodox” exercises of self-control, in which exercises of self-control are against the agent’s better judgement. Here, he discusses the cases of Bruce and Alex. Bruce, he explains, “has decided to join some friends in breaking into a neighbor’s house, even though he judges it best on the whole not to do so.”⁸¹ According to Mele, Bruce is afraid to perform the break-in and experiences “considerable trepidation” when he is about to do so. However, in spite of his fear, and importantly in spite of his better judgement, Bruce is able to master his fear, and thus exercise some form of self-control when he goes through with

⁷⁸ *Ibid*, pp. 45-46.

⁷⁹ *Ibid*, pp. 95-96.

⁸⁰ *Ibid*, p. 95.

⁸¹ *Ibid*, p. 94.

picking the lock of the house. Contra Frankfurt and Watson, who view unorthodox exercises of self-control as impossible, since self-control requires us to act on our better judgment on their views, in Mele's estimation, we can find an example in Bruce of the exercise of "self-control in the service of a decision that conflicts with his better judgment."⁸²

Perhaps the case of Alex makes Mele's point clearer. According to Mele's example, Bob's friend Alex "has proposed that they affirm their friendship by becoming blood brothers, because Alex is about to go away to prep school."⁸³ Becoming blood brothers involves the boys cutting each of their own hands and "shaking their hands so that the blood will mingle."⁸⁴ Alex critically reflects on his reasons for accepting the proposal, as well as his reasons for not accepting the proposal in terms of his aversion to cutting himself. Finally, Alex judges that it would be best to go ahead with the ceremony "at once," so that he might overcome his aversion. He forms the intention to "cut his hand with the knife straight away,"⁸⁵ and thus reaches for his knife moving it toward his hand "with the intention of drawing blood."⁸⁶ What Alex fails to do, however, is to factor in the possibility that he might find the task of actually cutting his own hand difficult to perform because of his aversion to cutting himself. Mele explains that as he gets the knife close to his skin, "he intentionally stops" his movement of the knife towards his hand. By stopping the movement, Alex does not "implement his original choice,"⁸⁷ but instead chooses to take the knife away and not perform the actions necessary for becoming blood brothers with Bob. In Mele's example, however, this does not mean that Alex changes his mind about what his better judgement on the matter is. That is, "he has not changed his mind about what it is best to do."⁸⁸ In fact, we could imagine that Alex is quite plausibly "upset with himself for chickening out."⁸⁹ What is made clear by the case examples of Bruce and Alex is that we can think quite easily of plausible cases in which an agent may exercise unorthodox self-control.

⁸² *Ibid*, p. 94.

⁸³ *Ibid*, p. 41.

⁸⁴ *Ibid*, p. 41.

⁸⁵ *Ibid*, p. 41.

⁸⁶ *Ibid*, p. 41.

⁸⁷ *Ibid*, p. 41.

⁸⁸ *Ibid*, p. 42.

⁸⁹ *Ibid*, p. 42.

That there are unorthodox exercises of self-control which are intentional or free (that is, which are also exercises of human agency) allows, I think, for the fact that while we may find ourselves acting intentionally against our better judgement, we still seem to experience our choices and actions in such cases as ‘plausibly our own’ or as ‘stemming just from ourselves.’ Not only does this fit with our intuitions about the bounds of our own agency, but points to a deeper problem with the view adopted by Frankfurt and Watson.⁹⁰ For Frankfurt and Watson, our better judgement constitutes the only legitimate aspect of ourselves, and all motivations or desires that run contrary to our better judgement are to be seen as ‘alien forces.’ Frankfurt says of such motives or desires that “its dominion is merely that of a tyrant. It has, for us, no legitimate authority.”⁹¹ Furthermore, he says:

Let us suppose that a certain motive has been rejected as unacceptable. Our attempt to immunize ourselves against it may not work. The resistance we mobilize may be insufficient. The externalized impulse or desire may succeed, by its sheer power, in defeating us and forcing its way. In that case, the outlaw imposes itself upon us without authority, and against our will.⁹²

Similarly, Watson says:

Here the opposition is not between you and another, but between you – that is, your evaluative judgment – and your other desires. Here, the “other” is your own motivation. This kind of conflict presents an issue of self-control rather than deliberation because here *insubordinate desires are to be resisted*. In these circumstances, their claims lack authority.⁹³

Although our better judgement plays a large role in the constitution of who we take ourselves to be, there are a number of other important motivations and desires that inform, in many basic ways, who we are. In comparison to Frankfurt and Watson, I think Mele rightly argues that “[e]ven when one’s passions and emotions run counter to one’s better judgment, they often are not plausibly seen

⁹⁰ I first discussed this idea in S. Paphitis, ‘Questions of the Self in the Personal Autonomy Debate: Some Critical Remarks on Frankfurt and Watson’, *The South African Journal of Philosophy*, vol. 29, No. 2, 2010, pp. 57-71.

⁹¹ H. Frankfurt, ‘Taking Ourselves Seriously’, p. 10.

⁹² *Ibid*, p. 14.

⁹³ G. Watson, *Agency and Answerability*, p. 62, emphasis added.

as alien forces.”⁹⁴ Take, for example, the case of Sue, who is a shopaholic despite her better judgement that it would be better to save for retirement and forgo needless purchases now. For Sue, shopping is an aspect of herself which informs her character and behaviour despite her better judgement. It is important to note here that the argument is not that we never act in accordance with our better judgement, or that our better judgement is not central to understanding human agency. Mele, rightly I think, is quick to point out that unorthodox cases of self-control will be less common than those of orthodox cases: people “who make attempts at self-control will tend to do so much more often in support of their better judgements than in opposition to them.”⁹⁵ The argument is that some motivations and desires that run counter to one’s better judgement can be integral aspects of agential life, even though, for the most part, we explain the self-control associated with human agency as guided by our better judgement. Mele says:

A conception of self-controlled individuals as, roughly, people who characteristically are guided by their better judgments even in the face of strong competing motivation does not commit one to viewing emotion, passion, and the like as having no place in the self of self-control.⁹⁶

We might support Frankfurt and Watson’s arguments here by suggesting that it is a normative demand that our actions should correspond with our better judgement. Even if, descriptively, we can exercise unorthodox self-control, such exercises are not agency conferring. In cases where we exercise unorthodox self-control we fail to do what good agents would do, and thus fail to be agents in Frankfurt and Watson’s sense. I think that this move is implausible for two related reasons: first, we would need further substantial argument that normative prescriptions for good agency should inform our descriptive accounts of human agency. Second, we would need further substantial argument that good agents are those who always act on their better judgement. These arguments are not given by Frankfurt and Watson, and in the face of our experience, as well as

⁹⁴ A. Mele, *Backsliding*, p. 92.

⁹⁵ *Ibid*, p. 94.

⁹⁶ *Ibid*, p. 92.

Mele's argument, I think that we should accept both orthodox and unorthodox exercises of self-control as exercises of human agency.⁹⁷

Further, that there are unorthodox exercises of self-control, which we associate with human agency, tells us that the kind of self-control central to agency fundamentally rests on a connection between, as Mele points out, "motivational strength and intentional action."⁹⁸ An agent exercises self-control when she is able to muster the motivational strength to perform intentional actions in the face of competing desires. This view of self-control which is centred on motivational strength allows us to make sense of cases in which a person exercises self-control in such a way that she acts contrary to her better judgement, and thus provides an important supplementation to Frankfurt' and Watson's accounts of self-control.

1.4 Internal Conflict

I argue in this section that the experience of conflicts between motivations and intentions on the part of a human agent is constitutive of self-control. That human agency requires internal conflict is an aspect of the self-control view which both Frankfurt and Watson to some extent discuss in their respective accounts. That internal conflict is required for agency, however, is an aspect of the self-control view of human agency which has, to some extent, been neglected in philosophical literature following the discussion and development of Frankfurt's and Watson's views. Drawing on both biological and psychological accounts of the nature of the mind, I defend in this section the claim that, on the self-control view, a human agent is someone who experiences conflict between her motivations and intentions. What makes the topic of human agency particularly interesting, is both the unique biological and distinctive psychological processes which underpin this complex human ability.

⁹⁷ It is important to note that an unorthodox exercise of self-control, that is to say acting against our all things considered better judgment, does not undermine our agency in the same way that being moved by forces such as 'tics' or neurosis usually would. The cases described in this section which outline the conditions for an agent to exercise unorthodox self-control differ from those which will be discussed in detail later in this thesis in which there is no exercise of self-control and in which agency is undermined. In cases where agency is undermined, as suggested here, there has been no motivational strength exercised on the part of the agent.

⁹⁸ *Ibid.*

I opened the first section of this chapter with a discussion of a biological picture of human agency suggested by Murphy and Brown. For Murphy and Brown the complexity of our neural systems allows us far more control over our behaviour qua comparators than the control expressed by any other organisms along the intentionality gradient. Following this view, I have endorsed in this chapter a philosophical account of agency centred on self-control, which stresses the centrality and importance of our ability to modify and guide our behaviour through critical reflection and assessment. As humans, we experience a very rich and complex mental life. Not only are we capable of having many disparate thoughts and feelings, but we are also capable of reflecting on them, deliberating on them, and evaluating them. Following the biological picture sketched by Murphy and Brown, what is inherent in the expanding complexity and range of available response behaviours in human beings is the potential for conflict between the behavioural response options. In turn, the potential for this conflict to further complicate the task of critical-reflection and evaluation undertaken by agents themselves (qua comparators in Murphy and Brown's terms) is greatly increased.

Similarly, on both Frankfurt and Watson's accounts, the presence of internal psychological conflict within an agent's volitional system is an important part of understanding the nature of human agency. Although this aspect of the self-control view of human agency has received little attention in the recent literature developed from Frankfurt's and Watson's accounts, I argue that it is an important feature of human agency on their accounts.

Inherent in the hierarchical structure of Frankfurt's account is the potential for conflict between desires of various orders. On Frankfurt's account, agents potentially experience not only conflict between their various first order desires, but also experience conflict between their various second order desires as well as between their first and second order desires. According to Frankfurt:

There is much opportunity for ambivalence, conflict and self-deception with regard to desires of the second order... as there is with regard to first-order desires.⁹⁹

⁹⁹ H. Frankfurt, 'Freedom of the Will and the Concept of a Person', p. 15.

To some extent similarly, for Watson not only is there potential for agents to experience conflict between their motivations and intentions, Watson argues that it is precisely the division between one's motivational system and evaluational system that gives rise to internal conflict. According to Watson:

It is an essential feature of the appetites and the passions that they engender (or consist in) desires whose existence and persistence are independent of the person's judgment of the good. The appetite of hunger involves a desire to eat which has a source in physical needs and physiological states of a hungry organism. And emotions such as anger and fear partly consist in spontaneous inclinations to do various things – to attack or flee the object of one's emotion, for example. It is intrinsic to the appetites and passions that appetitive and passionate beings can be motivated in spite of themselves. It is because desires such as these arise independently of the person's judgment and values [...] that a conflict between valuing and desiring is possible.¹⁰⁰

Psychologically speaking, not only are there tensions and conflicts between our brute or initial responses to the world, but also between the values and goals used in and for evaluation itself: we are what John Cottingham calls “conflicted beings” and “divided selves”.¹⁰¹ Our internal tensions and conflicts give rise to psychological divisions which when successfully negotiated can be resolved or managed. It is the successful resolution of internal conflicts which, following Frankfurt and Watson's accounts, allows for the exercise of the kind of self-control required for freedom of the will or free agency, or what I call human agency. Unlike Frankfurt and Watson, Cottingham points out, I think rightly, that in order to fully understand the role internal conflict plays in our psyche, and thus for our agency, we must turn to some of the crucial insights of psychoanalysis. Cottingham says:

The term ‘divided self’ is strongly suggestive of a psychoanalytic approach to understanding the human predicament; and my personal view is that this is precisely the direction we need to take, if we are tackling the problem of human conflictedness and its possible resolution [...] In short I think a full account [...] will sooner or later need to be receptive to some of the insights

¹⁰⁰ G. Watson, ‘Free Agency’, p. 21.

¹⁰¹ J. Cottingham, ‘Integrity and Fragmentation’, *Journal of Applied Philosophy*, vol. 27, no. 1, 2010, p. 6.

into the human condition articulated over the last century or so by Freud and Jung their successors.¹⁰²

In what follows I will show that the claims about the divided nature of the human psyche evident on Frankfurt's and Watson's accounts accord with a plausible basic account of human psychology following the influential structural account of our psyche given by the father of psychoanalysis, Sigmund Freud. I do not intend to defend a Freudian account in full here, but I will draw on some of the more foundational and less controversial aspects of Freud's most basic psychological account of the structure of the human psyche. Importantly, I take Freud's picture to mirror in important ways the pictures of the human psyche I have touched on in both the biological and philosophical accounts of human agency outlined and defended above.

On both the self-control view of agency established by Frankfurt and Watson and the basic Freudian account, the structure of an agent's mental apparatus is divided into a number of parts, each representative of different aspects of an agent's mental life. In order to exercise agency, then, the agent, on these pictures, must bring these sometimes conflicting parts together to form coherent motivations and intentions, which are in turn manifested in action and behaviour. For Frankfurt, our first order desires need to be brought in line with our second order volitions, and for Watson, our motivational and evaluative systems need to be brought into line with one another.

Similarly, Freud offers one of the most influential and, I think, nuanced psychological pictures of the structural features of the mind, or, as he calls it, an "individual's mental apparatus."¹⁰³ By studying the data obtained through his professional interactions with 'mental patients', Freud claimed that he was able to draw conclusions not only about those with psychological problems, but also about the structures and operations of a healthy, or normal agent's mind. According to Freud:

If we throw a crystal to the floor, it breaks but not into haphazard pieces. It comes apart along its lines of cleavage into fragments whose

¹⁰² *Ibid.*

¹⁰³ S. Freud, *Introductory Lectures on Psycho-Analysis*, Riviere, J. trans., New York, Columbia University Press, 1983, p. 90.

boundaries, though they were invisible, were predetermined by the crystal's structure. Mental patients are split and broken structures of the same kind.¹⁰⁴

This quotation illustrates one of the central tenets of his picture of our mental apparatus: within the structure of the mind, there are certain predetermined divisions. In other words, the mind is necessarily divided into a number of parts, each of which, for Freud, had very specific features and functions, and played a crucial role in an agent's psychological life. Throughout his writings, Freud was concerned with how each of these structures developed, and, more importantly, with the relationships and conflicts between them, as a human subject matured over time as a biological organism.

For Freud, our primary or brute responses to the world, which define and govern our experiences in infancy and early childhood, are to be found in the structure of the id, which he called "the oldest of mental provinces or agencies."¹⁰⁵ In this structure, Freud claimed, we would find an individual's reservoir of instincts, drives and wishes.¹⁰⁶

In contrast to the id, the ego is considered to develop and evolve in the course of maturation and growth through an individual's interaction with the external world in which she finds herself. While the id is definitive of the early stages in life, the balance of power begins to shift as the ego develops. As the new mental structure, the ego develops and begins to take a prominent role by defining and governing the experiences of an individual. The ego, then, takes on the role of mediating between the demands and responses of the id and the external world in which an individual finds herself.¹⁰⁷ According to Freud:

Under the influence of the real external world which surround us, one portion of the id has undergone a special development [...] a special organization has arisen which henceforward acts as an intermediary between the id and the external world. This region of our mental life has been given the name of *ego*.¹⁰⁸

¹⁰⁴ *Ibid*, p. 73.

¹⁰⁵ S. Freud, *An Outline of Psychoanalysis*, Strachey, J. trans., London, The Hogarth Press and the Institute of Psycho-analysis, 1949, p. 2.

¹⁰⁶ J. Sandler, and R. Wallerstein, *Freud's Models of the Mind: An Introduction*, Madison, International Universities Press, 1998, p. 168.

¹⁰⁷ *Ibid*, p. 180.

¹⁰⁸ S. Freud, *An Outline of Psychoanalysis*, p. 2.

As the ego develops and evolves, the id remains unchanged, acquiring no “ego-like functioning.” It is incapable of any reflection, deliberation, evaluation or value judgement.¹⁰⁹ In short, the id is incapable of anything other than seeking satisfaction or expression in action and behaviour. Freud says of the id:

We approach the id with analogies: we call it chaos, a cauldron full of seething excitations. We picture it as being open as its end to somatic influences, and as there taking up into itself instinctual needs which find their psychological expression in it [...]. It is filled with energy reaching it from the instincts, but it has no organization, produces no collective will, but only a striving to bring about the satisfaction of the instinctual needs [...].¹¹⁰

In contrast, the ego, on Freud’s picture is that specialised part of an agent’s mental apparatus which is capable of critical self-reflection and evaluation. Freud says that “[t]he ego can take itself as an object, can treat itself like other objects, can observe itself, criticize itself, and do Heaven knows what with itself.”¹¹¹ In turn, the ego takes on an active role in self-direction and is capable of controlling whether and how the demands of the id will be assimilated into coherent motivations and intentions in light of its interactions with the external world. So, much like the picture of self-control discussed in the previous section, Freud claims that there is a structure in the mind (the ego), which is both capable of higher order or evaluative thought, and the reflective assessment of the lower order or brute responses that are found in the mind. Concordant with the pictures of self-control discussed previously, Freud maintains that self-control, and thus our capacity to act as an agent, rests on the fact that the assessing structure of the ego has the power to prevent the lower order or brute responses (id in Freud’s terms) from finding expression in the agent’s actions and behaviours when the demands of the external world are taken into consideration. According to Freud:

As regards internal events, in relation to the id, it [the ego] performs that task by gaining control over the demands of the instincts, by

¹⁰⁹ J. Sandler, and R. Wallerstein, *Freud’s Models of the Mind*, p. 178.

¹¹⁰ S. Freud, *Introductory Lectures on Psycho-Analysis*, pp. 91-92.

¹¹¹ S. Freud, *An Outline of Psychoanalysis*, pp. 77-78.

deciding whether they shall be allowed to obtain satisfaction, by postponing that satisfaction to times and circumstances favourable in the external world or by suppressing their excitations completely.¹¹²

For Freud, in order to become an agent the contents of the id must be given some life or expression. After all, for him, the id itself provides the psychic energy necessary for action.¹¹³ However, in order to exercise self-control the contents of the id must be given expression through their assimilation into a coherent set of intentions for action and behaviour, as mediated by the ego. In this way, much like for Frankfurt and Watson, while Freud sees the ego as representative of our better judgement, he sees the id as representative of our brute responses. Freud claims:

The ego develops from perceiving the instincts to controlling them; but this last is only achieved by the [psychical] representative of the instinct being allotted its proper place in a considerable assemblage, by its being taken up into a coherent context. To adopt a popular mode of speaking, we might say that the ego stands for reason and good sense while the id stands for the untamed passions.¹¹⁴

The source of energy for action, which we might call motivation, is located in a part of our primal, instinctive and chaotic psyche. The energy given by the id must be channelled and directed by the ego, the task of which is to mediate between the conflicting demands of the id and the external world. On the Freudian account, exercising human agency rests on our being able to reconcile the demands of competing psychical forces:

Thus, an action by the ego is as it should be if it satisfies simultaneously the demands of the id [...] and of reality, that is to say if it is able to reconcile their demands with one another.¹¹⁵

Like Plato's analogy of the charioteer, in which the relationship between what he calls the various parts of the human soul is compared to the relationship between a charioteer and his horses, Freud compares the relationship between

¹¹² *Ibid*, p. 3.

¹¹³ Freud talks about the id as 'driving' the ego; and talks about the ego as needing to attract the 'libido' of the id for momentum. S. Freud, *Introductory Lectures on Psycho-Analysis*, pp 97-98.

¹¹⁴ *Ibid*, p. 95.

¹¹⁵ S. Freud, *An Outline of Psychoanalysis*, pp. 3-4.

the ego and the id to the relationship between a horse and its rider. The ego, like both the rider and Murphy and Nancy's comparator, is capable of control by directing and guiding the powerful locomotive force beneath it. The id, however, or the "seething cauldron" of conflict, is at times too powerful and strong-willed for the ego to control. According to Freud:

The ego must on the whole carry out the id's intentions, it fulfils its task by finding out the circumstances in which those intentions can best be achieved. The ego's relation to the id might be compared with that of a rider to his horse. The horse supplies the locomotive energy, while the rider has the privilege of deciding on the goal and of guiding the powerful animal's movement. But only too often there arises between the ego and the id the not precisely ideal situation of the rider being obliged to guide the horse along the path by which it itself wants to go.¹¹⁶

On Freud's account, then, conflict within the psyche exists necessarily because of inconsistencies and tensions within the id itself. The id, in which contrary impulses exist side by side, is a structure in the mind which is not governed by reason. Freud says:

The logical laws of thought do not apply in the id, and this is true above all of the law of contradiction. Contrary impulses exist side by side, without cancelling each other out or diminishing each other [...].¹¹⁷

The ego must thus, in addition to harmonising the demands of the id with the external world, harmonise the competing demands found in the id itself. Furthermore, the psychological conflict is made difficult for an agent to manage because the ego has an allegiance to both the id and the external world. Freud says of the ego's situation:

Owing to its origin from the experiences of the perceptual system, it is earmarked for representing the demands of the external world, but it strives too to be a loyal servant of the id, to remain on good terms with it, to recommend itself to it as an object and to attract its libido to itself.¹¹⁸

¹¹⁶ S. Freud, *Introductory Lectures on Psycho-Analysis*, p. 96.

¹¹⁷ *Ibid*, p. 92.

¹¹⁸ *Ibid*, pp. 97-98.

The ego, we might say, is caught between a rock and a hard place: pressed between a demanding id, the locomotive energy of which is required for it to function, and the demands of an external world in which this locomotive energy must be discharged.

Since each of the structures described by Freud provides us with possible motivations, the conflicts between the motivations which arise within these separate structures explain, in psychological terms, the possibility of internal conflict for human agents. We have all experienced being simply overwhelmed by some desire or motivation which prevents us from being able to exercise self-control: we have all been overcome by the temptations we feel the pull of most strongly, giving in to the desires which are most strongly rejected by our better judgement, and finding ourselves from time to time wholly lacking in self-control or without any restraint. Most importantly, an interesting conclusion of Freud's account is that the activity associated with agency seems to require that we experience conflict and constraint between our motivations. Indeed, without this internal conflict it is unclear that we would need to exercise the motivational strength central to self-control at all.

Similarly, on biological grounds, or at least from the naturalist's perspective, Mary Midgley has championed a view of human agency in which both self-control and internal conflict play a central role. According to her, self-control or behavioural modification is not simply a structural feature of human agency, but rather the central activity in which we must actively engage in order to exercise our agency.¹¹⁹ For Midgley, the kind of mental or psychological activity associated with self-control is what she refers to as the "organisation of the inner crowd,"¹²⁰ by which she means the ability to manage the conflicts that occur within our minds (such as conflicts between more than one desire, between desires and intentions and between intentions themselves).¹²¹ For her human agency requires an agent to experience psychological divisions and tensions, because, without them, there is no 'inner crowd' to control, and hence

¹¹⁹ M. Midgley, *Beast and Man: The Roots of Human Nature*, London, Methuen, 1979. M. Midgley, *The Ethical Primate: Humans, Freedom, Morality*, London, Routledge, 1994. M. Midgley, *Wickedness: A Philosophical Essay*, London, Routledge, 2000.

¹²⁰ M. Midgley, *The Ethical Primate*, p. 183.

¹²¹ I will return to conflicts between intentions in Chapter 4 in more detail.

the possibility for self-control itself cannot get off the ground. The kind of self-control that I have argued is constitutive of human agency must involve, then, most fundamentally a response to internal division and constraint, making internal conflict central to understanding the kind of psychological activity we would associate with human agency. Because of this internal conflict and division, a large part of our mental activity qua agents requires, first and foremost, organisation and conflict resolution. Without our efforts to resolve these types of conflict, it would be impossible for us to successfully modify and direct our behaviour in the face of all the different desires, emotions, values and loyalties which may divide us.

Importantly, I think that it would be a mistake to think of internal conflict as a merely contingent aspect of human agency. In what follows then I argue for an addition to the core of the self-control view of human agency as follows: human agency requires at the very least the possibility of a discrepancy between motivations and evaluations, and, in turn, the experience of internal conflict on the part of an agent. I argue here that on the self-control view of human agency: an agent is to be understood as someone who experiences conflict between motivations and intentions.

For Watson there is a deep connection between self-control and internal conflict, and, for him, the possibility of internal conflict within our motivational structure is in fact necessary for the evaluative act of self-control, he argues:

Thus, the relationship between evaluation and motivation is intricate. With respect to many of our activities, evaluation depends upon the possibility of our being moved to act independently of our judgment.¹²²

On Watson's account, and similarly on Frankfurt's and Midgley's accounts, internal conflict is not only implicated in the capacity for self-control, it is in fact, as I see it, necessary for it. An individual could not exercise the capacity for self-control at all unless she has something which needs to be controlled. Not only is internal conflict part of the biological and psychological story we can tell about

¹²² G. Watson, 'Free Agency', p. 23.

the nature of human agency, it is a constitutive part of human agency on a philosophical account as well.

Freud was most interested in those deep, unyielding and constant internal conflicts which produced in his patients a psychic world in which the self-control required for human agency was no longer attainable. The psychoanalytic approach defended by Freud is intended, at least in part, to address the issue of overwhelming internal conflict in order to restore healthy psychic harmony for the patient to be able to regain their agency through a measure of self-control – or, as Midgley might put it, to bring patients to a point of recovery from which they could start to effectively manage the inner crowd. But just as surely as too much internal conflict destroys our capacity for self-control, a psyche in which no internal conflicts are present at all necessarily precludes the possibility of self-control, and thus of human agency. In this way, internal conflict marks human agents out as distinct from beings, perhaps like that of a god or idealised ‘saint’, whose psychic life does not resemble ours, and whose agency does not most fundamentally require self-control.

In the psychic life of a god there may be no conflicts between their motivations, intentions, goals, aims and desires. Certainly, for God (as he is traditionally conceived, in the monotheistic traditions, as omnipotent and omniscient) there is no disparity between his motivational system and his evaluational system. God’s “dependence of motivation upon evaluation is total, for there is but a single source of motivation: presumably his infinite goodness and wisdom, whatever that may mean.”¹²³

Similarly, in the case of a saintly being, who has the requisite internal division but experiences no such conflicts, she may sometimes find herself with motivations which run counter to her intentions; yet she has no *struggle* between her wayward motivations and her intentions when it comes to acting, that is, she has no internal conflict. While she acts as the self-controlled person would, her actions do not require exerting self-control at all or ever. She is a being, I think, who ought to be distinguished from human agents like ourselves. For a human agent, while there may be particular cases in which they are not conflicted (or are at least not struggling with internal conflict), we might still

¹²³ *Ibid*, p. 31.

want to say that they retain self-control. As a general capacity the self-control necessary for human agency cannot be had without the very real possibility of internal conflict. This is an aspect of our agency which we share with neither non-human animals, nor beings perhaps like that of a god or an idealised 'saint.' I will return to this claim in the next chapter for further discussion.

Chapter Two: Vulnerability and the Difficulty of Self-Control

Mortals immortal, immortals mortal, alive with respect to mortals' death, dead with respect to their life.¹²⁴

The first humans, as Lucretius describes them, found life sweet. They left the "sweet shores" of life with sadness and looked to their departure with fear (V.988-93). The love of life, Lucretius claims, is natural in all sentient creatures, and so all creatures go to death with reluctance. But these "first men" do not stop [...] to reflect on their finitude. They do not wonder about their own fragility, or find agony in the mere knowledge of the "mortality of life." The Epicurean gods, on the other side, have (as the poem describes them) reflection without vulnerability, thought about the universe without anxious fear and concern. *In between are actual human beings, the only beings both vulnerable and reflective, who go through life in the grip of a fear of the natural condition of their own existence, straining to understand and also to improve their condition through the reflective capacity that is also the source of much of their agony.*¹²⁵

In the second epigraph to this chapter, drawn from Martha Nussbaum's paper 'Mortal Immortals', Nussbaum defends the view that our vulnerability is essential for understanding our situation in the world, our experience as agents in the world, and what constitutes value from our perspective. In that paper, however, Nussbaum explicitly defends at length only two aspects of this view, namely that vulnerability is essential for understanding our situation in the world and that vulnerability is central to our understanding of what value consists in for us.

Nussbaum argues in 'Mortal Immortals' that "the structure of human experience, and therefore of the empirical human sense of value, is inseparable from the finite and temporal structure within which human life is actually lived."¹²⁶ She argues that it is "[o]ur finitude, and in particular our mortality, which is a particularly central case of our finitude, and which conditions all our other awareness of limit, [that] is a constitutive factor in all valuable things' having for us the value that they in fact have."¹²⁷ It is this finitude, made most poignantly manifest in our mortality, that is also a source of a deeper

¹²⁴ Heraclitus, as quoted in M. Nussbaum, 'Mortal Immortals: Lucretius on Death and the Voice of Nature', *Philosophy and Phenomenological Research*, vol. 50, no. 2, 1989, p. 303.

¹²⁵ *Ibid*, pp. 304-305, my emphasis.

¹²⁶ *Ibid*, p. 336.

¹²⁷ *Ibid*.

vulnerability. This vulnerability is not only contingently necessary for the emergence of human experience as we know it, but it is constitutive of that experience. Further, she claims that it is this vulnerability which frames our understanding of what the valuable things are for us, arguing that “the removal of all limit, of all constraint of finitude in general, mortality in particular, would not so much enable these values to survive eternally as bring about the death of all value as we know it.”¹²⁸ She argues that it is precisely our vulnerability which distinguishes our lived experiences, and thus the determination of our values, from that of immortal and invulnerable beings such as the ancient Greek gods. That human vulnerability is central to our understanding of ourselves and of our values in this way is movingly motivated for by Nussbaum when she says:

In these constraints we live, and see whatever we see, cherish whatever we cherish, as beings moving in the way we actually move, from birth through time to a necessary death. The activities we love and cherish would not, as such, be available to a godlike unlimited being.¹²⁹

Central to this line of argument, for Nussbaum, is the importance of the empirical nature of our understanding of value, and thus Nussbaum claims that any exploration following this line of argument “will have to be empirical, bringing forward parts of human experience and using them (rather than some alleged *a priori* principles) to establish that our concept of value is in fact the way we say it is.”¹³⁰ Further, and importantly, she says that this kind of phenomenological or empirical exploration “will also need to be deep and probing, going beneath first impressions and automatic everyday responses, to elicit our deepest judgments about what matters.”¹³¹ Following Nussbaum’s recommendation here, as well as the broader line of argument brought out in the second epigraph to this chapter – namely, that vulnerability is central to our understanding of our agency in the world – I discuss, in this chapter, what can be seen as contingently necessary requirements for human agency in terms of vulnerability and difficulty. Further, again following Nussbaum’s lead, I argue that difficulty and vulnerability are not only background causal requirements for agency but are constitutive of human

¹²⁸ *Ibid.*

¹²⁹ *Ibid.*

¹³⁰ *Ibid.*, p. 336.

¹³¹ *Ibid.*, pp. 336-337.

agency. It is precisely this requirement for difficulty and vulnerability that distinguishes our unique human agency from that of both other living beings, and invulnerable, unyielding beings such as the gods.

In the previous chapter, I outlined and defended a self-control view of human agency. While a great deal of debate has been generated by the self-control picture of agency, somewhat less attention has been paid to the more practical issues involved in actually exercising agency of this kind. Of course, theoretical pictures are useful and interesting in their own right, but an important added bonus would be if they helped us to better understand our practical situation. When it comes to practical situations, robust and interesting pictures of agency such as those provided by Frankfurt and Watson, provide little insight into the difficulty that is personally experienced by each of us when we attempt to exercise the kind of self-control they describe. Practically speaking, there are a very great number of obstacles we face which make it difficult for us to perform the task of effectively using the product of our better judgement or reflective evaluations to guide our actions and behaviour. These obstacles are not limited, as the philosophical literature often seems to suggest, to the bizarre cases, cooked up in thought experiments, of brainwashing and mental manipulation. Rather, they are present in everyday experiences that human agents face.

In the previous chapter, I mentioned that one extreme situation in which such constraints might be observed is within the walls of a concentration camp. Recall here Frankl's extreme, and I think incorrect view, on which even in situations of extreme constraint, such as those in Auschwitz, we should not see human agency as having been undermined by the constraints people would have faced there. According to Frankl:

Even though conditions such as lack of sleep, insufficient food and various mental stresses may suggest that the inmates were bound to react in certain ways, in the final analysis it becomes clear that the sort of person the prisoner became was the result of an inner decision, and not the result of camp influences alone. Fundamentally, therefore, any man can, even under such circumstances, decide what shall become of him – mentally and spiritually.¹³²

¹³² V. Frankl, *Man's Search for Meaning*, p. 74.

While Frankl briefly touches on the idea that it would, in fact, be very difficult to maintain one's inner liberty in the environment of a concentration camp, he goes on, I think erroneously, to suggest that while it would be difficult, it is something that any of us could do, practically speaking. He writes:

It is true that only a few people are capable of reaching such high moral standards. Of the prisoners only a few kept their full inner liberty and obtained those values which their suffering afforded, but even one such example is sufficient proof that man's inner strength may raise him above his outward fate.¹³³

Frankl suggests that although the psychological and physical influences of our living conditions can rob us of our inner freedom, we can also overcome these influences and retain our agency in the face of them. In this chapter, I explore in more depth Frankl's basic idea that agency may be difficult to exercise under certain conditions. I argue, however contra Frankl, that it is far more commonplace than he suggests that the circumstances in which we find ourselves place us under such an enormous amount of pressure to act and behave in certain ways, that it becomes very difficult or ultimately impossible to truly exercise the kind of self-control required for human agency. Experiencing hunger, fear, lack of sleep or pain, is often the source of unreflective behaviour, over which we experience little control. Moreover, having to exercise such vigilant self-control in these kinds of situations can also diminish our willpower,¹³⁴ resulting in the kind of behaviour which goes against our own better judgements or intentions. When we are able to exercise self-control, in the face of constraints of this type, we are aware of the immense effort involved, which often leaves us feeling emotionally, psychologically and sometimes even physically drained, and therefore less capable of exercising self-control.

¹³³ Frankl is also incorrect here, I think, that we could generalize from one exemplary case. V. Frankl, *Man's Search for Meaning*, p. 76.

¹³⁴ When our willpower is diminished it is more difficult to exercise the kind of self-control necessary for agency, see experiments on decision fatigue in D. Ariely and G. Loewenstein, 'When does Duration Matter in Judgment and Decision Making?', *Journal of Experimental Psychology: General*, vol. 129, 2000, pp. 508-29. D. Ariely, and D. Zakay, 'A Timely Account of the Role of Duration in Decision Making', *Acta Psychologica*, vol. 108, no. 2, 2001, pp. 187-207. G. Loewenstein 'Willpower: A Decision-Theorist's Perspective', *Law and Philosophy*, vol. 19, 2000, pp. 51-76.

Here I argue, contra Frankl, that recent work in the fields of behavioural economics and experimental psychology provide us with everyday examples in which people act against their own better judgements or critical evaluations without even realising that their agency has been diminished by their situation. More importantly perhaps, such work provides us with further examples of everyday situations in which our better judgements or intentions are themselves systematically tainted by the influence of external forces of which we are largely unaware. In the second chapter of this thesis, then, I highlight the difficulty of achieving the kind of self-control required for human agency on formal analytic accounts by revealing the extent and systematic depth of the kinds of constraints we in fact face. I think that the difficulty we face in attempting to exercise self-control allows us to make a crucial observation about the nature of human agency which is not yet fully captured in formal analytic accounts, such as those discussed in the first chapter: namely that human agency is difficult to achieve on account of our vulnerability. I argue that the kind of self-control necessary for human agency is difficult to achieve because we are vulnerable in a number of distinct ways. First, our human bodies are vulnerable, and because we are embodied, self-control is difficult to achieve. Second, we are vulnerable to situational constraints, and because we are in large part constituted by our situation, self-control is difficult to achieve. Third, we are cognitively vulnerable, and since we employ flawed heuristics and have cognitive biases, self-control is difficult to achieve. Finally, as divided beings we are subject to internal vulnerability, because we experience internal conflict between motivations and intentions, self-control is difficult to achieve.

If we are vulnerable, it is important to ask if there is anything we can do about this vulnerability: might there be ways to improve our chances of self-control, improve our agency, and thus extend our freedom? I will argue that we can answer this question, to some extent, in the affirmative, and thus we can make some practical suggestions for improving our self-control. However, as suggested above, I maintain that vulnerability is constitutive of human agency, and thus any attempt to wholly eradicate vulnerability would also serve, on my view, to eradicate human agency.

In defence of Nussbaum's view expressed in the second epigraph to this chapter that "human beings [are] the only beings both vulnerable and reflective, who go through life in the grip of a fear of the natural condition of their own existence, straining to understand and also to improve their condition through the reflective capacity that is also the source of much of their agony", I argue in what follows that we can gain valuable insight into the nature of human agency from the existentialist philosophers Friedrich Nietzsche and Simone de Beauvoir, both of whom explore the psychology and phenomenology of agency respectively. Following the insights from both Nietzsche and Beauvoir, as well as the evidence from experimental psychology and behavioural economics, I argue that on the self-control view of human agency, both self-control and vulnerability are necessary for human agency. I argue here that we must supplement the self-control view of human agency, as discussed in the first chapter, in an important way: human agency should be seen as difficult to achieve on account of our vulnerability.

2.1 Vulnerability and Difficulty

What is called "freedom of the will" is essentially the affect of superiority with respect to something that must obey: "I am free, 'it' must obey" – this consciousness lies in every will, along with a certain straining of attention, a straight look that fixes on one thing and one thing only, an unconditional evaluation "now this is necessary and nothing else," an inner certainty that it will be obeyed, and whatever else comes with the position of the commander. A person who wills –, commands something inside himself that obeys, or that he believes to obey. But now we notice the strangest thing about the will – about this multifarious thing that people have only one word for. On the one hand, we are, under the circumstances, both the one who commands and the one who obeys, and as the obedient one we are familiar with the feelings of compulsion, force, pressure, resistance, and motion that generally start right after the act of willing. On the other hand, however, we are in the habit of ignoring and deceiving ourselves about this duality by means of the synthetic concept of the "I." [...] "Freedom of the will" – that is the word for the multi-faceted state of pleasure of one who commands and, at the same time, identifies himself with the accomplished act of willing. As such, he enjoys the triumph over resistances, but thinks to himself that it was his will alone that truly overcame the resistance. Accordingly, the one who wills takes his feeling of pleasure as the commander, and adds to it the feelings of pleasure from the successful instruments that carry out the task, as well as from the useful "under-wills" or under-souls – our body is, after all, only a society constructed out of many souls –. *L'effet c'est moi*: (The effect is I) what

happens here is what happens in every well-constructed and happy community: the ruling class identifies itself with the successes of the community.¹³⁵

In my introduction I discussed that my engagement with the writings of Harry Frankfurt which had first drawn my attention to the nature of the relationship between human agency and personhood, but there is more to this story. While reading the writings of Harry Frankfurt I found myself simultaneously reading Friedrich Nietzsche. Perhaps it was merely this parallel reading of these two influential philosophers that alerted my attention to a common ground they seem to share in their discussions of the nature of human agency and personhood. In this section, I will argue that that the account of human agency developed in Nietzsche's writings, at least according to one plausible reading of Nietzsche, accords in many ways with the self-control view of agency built on the basic Frankfurtian framework as outlined in the first chapter.

In what follows in this section, I highlight two of Nietzsche's suggestions about the nature of human agency which lack clarity and exposition in Frankfurt's account (and similarly in Watson's account, as well as much of the literature on human agency developed in response to their accounts). I defend throughout this chapter Nietzsche's suggestions that human agents are vulnerable and that human agency is difficult to achieve. In drawing on Nietzsche's works here I rely to a large extent on my own reading of Nietzsche, which I have found support for in recent Nietzsche scholarship. While some of Nietzsche's more controversial claims might have put many readers permanently off Nietzsche's philosophy, I think that there are important claims (particularly about the nature of human agency) made by Nietzsche which nevertheless warrant our attention.

Although Nietzsche's sceptical views on freedom have been seen as controversial in many philosophical circles, it has recently and convincingly been argued by a number of philosophers¹³⁶ that Nietzsche does advance a positive account of freedom or an idea of agency in his discussion of his character the Sovereign Individual, as well as in his concept of self-overcoming. In this section I

¹³⁵ BGE 1: 19.

¹³⁶ See the collection in K. Gemes, and S. May, (eds.) *Nietzsche on Freedom and Autonomy*, Oxford, Oxford University Press, 2009.

show that Nietzsche asserts, in part through his characterisation of the Sovereign Individual, that some form of self-control is required for the project of becoming an agent. It is important to note before delving into Nietzsche's account that reference to the Sovereign Individual is only explicitly made by Nietzsche in *On the Genealogy of Morals* Section II, and might thus not represent the only picture of agency which could be drawn from Nietzsche's writings. In this section, and the sections which make reference to the Sovereign Individual that follow in this thesis, I have drawn on his conception of the Sovereign Individual in the *Genealogy*, but have tried to put forward a picture of the Sovereign Individual which draws on ideas and claims from other parts of Nietzsche's work for supplementation and explanation. Further, while I realise that it would take a great deal of interpretive argument (for which there is not much room here) to claim that there is a definitive and explicit notion of agency in Nietzsche's work, this project is in part an attempt to tease out at least one plausible reading of Nietzsche's thoughts about this concept. This is not the place to deal in depth with exegetical matters. Rather, what I will do here is defend an account of human agency which was fuelled by my readings of Nietzsche. Whether my interpretation of Nietzsche is actually correct is not my central concern here.

Similarly to the self-control view of agency sketched in the previous chapter, Nietzsche emphasises self-governance or self-control in both the motivational and evaluative senses. This is made "particularly prominent in [his] later works like *Twilight of the Idols*."¹³⁷ Much like the agent on the self-control view sketched in the first chapter, Sovereign Individuals, on Nietzsche's account, are thought of as actively asserting control over or governing themselves from the inside, thus mastering conflicting inclinations and motivations. For Nietzsche, agents form values on the basis of their brute desires. But these values are not simply formed on the basis of *all* of these desires. This process involves, for the agent, selecting and affirming or endorsing some desires over others with which they may be in conflict. This process also resembles the picture of agency given in analytic accounts, as sketched above, in which our intentions guide our deliberations

¹³⁷ R. Pippin, 'How to Overcome Oneself: Nietzsche on Freedom', in Gemes, K. and May, S., (eds.) *Nietzsche on Freedom and Autonomy*, Oxford, Oxford University Press, 2009, p. 76.

about what motivations to endorse and act upon. For Nietzsche, mastering or controlling our conflicting motivations in the service of the values we endorse is a fundamental part of exercising our agency. Nietzsche explains as follows:

Indeed, where the plant 'man' shows himself strongest one finds instincts that conflict powerfully [...] but are controlled.¹³⁸

It is our ability to master and control conflicting desires that, for Nietzsche, similarly to Midgley, most fundamentally represents our ability to overcome ourselves: in order to follow through on our intentions, we must overcome those conflicting desires and inclinations that would otherwise motivate us to act against our intentions, which are also, importantly, our own (this is what Nietzsche refers to as self-overcoming). In Nietzsche's view, if we are not able to control our inner conflicts (at least some of the time), we are not capable of exercising our agency or becoming Sovereign Individuals. In agreement with Gemes' recent discussion of Nietzsche on agency, I argue that if we are not able to exercise our agency, it is a most dangerous threat to our sovereignty because it undermines our right to make promises, something which is perhaps the defining characteristic of the Sovereign Individual *qua* agent. In *The Genealogy of Morals*, Nietzsche says:

We discover that the ripest fruit is the sovereign individual, like only to himself [...] autonomous and supramoral [...] [I]n short, the man who has his own independent, protracted will and the right to make promises – and in him a proud consciousness [...] of his own power and freedom, a sensation of mankind come to completion. This emancipated individual, with the actual right to make promises, this master of free will, this sovereign man.¹³⁹

For Nietzsche, the Sovereign Individual has the right to make promises precisely because he is able to exercise the kind of self-control central to the self-control view of agency outlined in chapter one. Gemes explains, that for Nietzsche, one cannot have agency in any genuine sense if one is “merely tossed about willy-

¹³⁸ WP 966.

¹³⁹ GM II: 2.

nilly by a jumble of competing desires.”¹⁴⁰ For Nietzsche, unless you are able to exercise control over yourself, “you cannot stand surety for what you promise.”¹⁴¹ If you are unable to master your conflicting motivations, you cannot be sure that you will honour your promise when it comes to acting on it, since you may well act on a conflicting or contrary inclination at any time. If you cannot stand surety for your promises, Nietzsche thinks that you have not earned the right to make promises at all. Furthermore, unless you have the right to make promises, you cannot be an agent or, in Nietzsche’s terms, a Sovereign Individual.

It is, of course, interesting to note the similarities between Nietzsche’s sometimes controversial account of agency and the accounts proposed more recently by analytic philosophers. There is, however, something particular that comes out of Nietzsche’s account about the difficulty of motivational self-control, which I think requires closer examination. For Nietzsche, the difficulty involved in exercising self-control reveals the essential vulnerability of our agency. Nietzsche’s suggestion here is intuitively plausible. The difficulty of fighting inclinations, which go against our own better judgement, is something that is close to home for most of us, and something that we understand from our own experience as agents. We have all been overwhelmed by the temptations that we feel the pull of most strongly, given in to the desires which are most strongly rejected by our better judgement, and found ourselves from time to time wholly lacking in self-control or without any restraint. When our intentions are informed by our better judgement, then, it will be particularly difficult for us to act intentionally.

Following Nietzsche’s lead, and the lead of existentialists such as Beauvoir, I will argue in this chapter that self-control is difficult for us to achieve because we are vulnerable; and yet, we can only have self-control because we are vulnerable.

Typically, vulnerability refers to the inability of something to hold up under the effects of an adverse environment - someone is vulnerable because they are threatened, and thereby susceptible to harm. Human agency is

¹⁴⁰ K. Gemes , ‘Nietzsche on Free Will, Autonomy and the Sovereign Individual’, in Gemes, K. and May, S. (eds.), *Nietzsche on Freedom and Autonomy*, Oxford, Oxford University Press, 2009, p. 37.

¹⁴¹ *Ibid.*

vulnerable because it is susceptible to damage, compromise or dissolution and I will argue in what follows that human agents are vulnerable in a number of different, though interrelated, ways. First, I argue that because human agents are embodied, the fact of their embodiment makes their agency difficult to achieve and that human agency is vulnerable as a result. Second, I argue that the situations in which agents find themselves can make self-control difficult to achieve. Human agency thus comprises what I call situational vulnerability. Third, I argue that human agents are prone to cognitive constraints, which puts their ability to exercise self-control at risk, thus making human agency vulnerable. Finally, I argue that because human agents are subject to internal conflict, self-control is difficult to achieve, and thus human agency comprises what I call internal vulnerability.

Importantly, my discussion of vulnerability here should not be taken to mean that we are in any way defective because we are vulnerable. On the contrary, I argue throughout this thesis that vulnerability should be seen as both an obstacle to our agency, as well as, and importantly, the condition for the possibility of human agency.

In what follows in this chapter, I will outline what I take to be our primary vulnerabilities, and identify the ways in which these vulnerabilities put our agency at risk. I will then argue that these risks to our agency can, to some extent, be mitigated, but that on the self-control view, human agency must nevertheless be understood as both difficult to achieve and as essentially vulnerable.

2.2 Self-Control and Embodiment

In the previous chapter, I argued that the inclusion of internal conflict in the self-control view of human agency allows us to distinguish between the different kinds of agency we would consider animals, supernatural beings and human beings to have. I argued above that the philosophically interesting concept of human agency is at stake in the self-control view. That is, the notion of human agency applies only to human beings, as distinct from non-human animals and non-human supernatural beings. In line with these arguments, I suggest in this

section that our necessary embodiment is an important part of the story we can tell about human agency.

The claim that embodiment is crucial for understanding human agency might attract some initial criticism. It might, for instance, be argued by some that what in fact most interestingly defines us as human beings is our remarkable mental life – that is everything but the body. Recall here Frankl’s talk of the inner freedom he takes to be central to human agency. I do not wish to bring into question the importance of the kind of inner freedom associated with our mental life, which Frankl discusses (and which has largely been the focus of my discussion on agency in the previous chapter). Instead, I stress in this section the equally important role that embodiment plays in our understanding of what a human is qua human being, and thus I stress the fact that our embodiment is important for understanding the nature of *human* agency. The fact of our embodiment has most influentially been discussed by Beauvoir in her phenomenological account of the human condition. Given this, I draw inspiration from her phenomenology to explore the extent to which embodiment is a fundamental aspect of human agency. Following Beauvoir’s insights, contra Frankl, our bodies cannot be seen as simply a mode of transportation for our mental qualities, rather the fact of our embodiment is intricately linked to those mental qualities and is equally integral to our experience of ourselves as human agents. Following this, I argue that it is because of human embodiment that we are subject to what I call bodily vulnerability. Our bodies themselves are vulnerable things: they are prone to injuries and sickness, grow old and die. Because human agents are embodied, I suggest, their ability to exercise self-control is constantly at risk and the self-control required for human agency is difficult to achieve. We are vulnerable because we are what Beauvoir explains as ‘embodied subjects’: we are inextricably linked to our body, and our body is vulnerable.

In breaking with the traditional Cartesian picture suggested by Frankl, existential phenomenologists – such as Beauvoir – take embodiment as central to our self-constitution, and the fact of our embodiment as crucial for understanding our experience of ourselves as human subjects. For the existential phenomenologists,

our unique way of being in the world is explained by our mental qualities and capacities, *as well as embodiment*, which is fundamentally integral to this situation. We live as embodied agents and so our actions in the world are, to some extent, determined by our bodies in terms of their functions, manifestations and limitations. Our being embodied allows us to act in the world because it is through our bodies that we can interact with objects in the world: our bodies enable action and thus, to some extent, being embodied allows for the very possibility of human agency in the first place. For Beauvoir, being embodied enables us to express our mental life (comprising our values, projects and goals) in the world (to which the construction of our mental lives is a response; and in which our goals and projects must be acted out). To this extent, embodiment is crucial to our agency on a phenomenological picture.

Further, the body, for Beauvoir, is not merely a thing like other objects in the world. For us, according to Beauvoir, the body 'is a situation in its own right'. Beauvoir explains:

In the perspective I am adopting—that of Heidegger, Sartre, and Merleau-Ponty—if the body is not a thing, it is a situation: it is our grasp upon the world and the outline of our projects.¹⁴²

The body is not an object in the world, but positions us – as embodied subjects – in the world. While our embodiment allows us to express ourselves in action in the world, it also presents us with certain material facts about ourselves that directly constrain the types of actions we are able to perform.

Some of our bodily functions are not part of our conscious experiences — digestion and the circulation of blood happen, for the most part, behind the veils. Typically, we are made aware of some of these otherwise hidden bodily functions when things go wrong, as in when we experience the pain of indigestion or suffer from an illness. Other bodily functions, however, constitute a significant aspect of our embodied situation, and thus inform, in part, our lived experience as embodied subjects. That we must eat, sleep, urinate and defecate, and that we are prone to illness, harm and death, are facts about our general

¹⁴² S. Beauvoir, (DSI 78 SS 69), in S. Heinämaa, *Toward a Phenomenology of Sexual Difference*, United States of America, Rowman and Littlefield Publishers, Inc., 2003, p. 25.

embodied situation that we must manage and accommodate in our day to day lives. We must factor them in to the formation of the projects and goals that inform how we choose to live. We are all well aware that we cannot escape having bodies and thus bodily functions. But because we do not always have full control over our bodily functions, we experience a certain sense of anxiety with regards to our bodies. We are anxious, I think, because we recognise our bodily vulnerability, and we see our bodily vulnerability as a threat to our agency.

Because our bodies require certain things to survive, and because our bodies can be damaged, we are made vulnerable to the extent that we can both be deprived of our bodily needs and be injured. It is because of our bodies that we experience fear, which is perhaps one of the strongest motivations experienced by all living, embodied things. Threats of deprivation, harm and death are more important considerations and stronger motivators for people than is suggested by authors like Frankl, and, similarly, philosophers such as Frankfurt and Watson.¹⁴³ Because our bodies are so easily threatened, our agency is also threatened precisely because we are more susceptible to tactics of coercion and manipulation which centre on bodily harm and deprivation. Experiencing hunger, fear, lack of sleep, or pain, is often the source of unreflective behaviour. Extreme situations, like being imprisoned in a concentration camp may highlight the difficulty most clearly. Even Frankl admits that under such extreme circumstances the examples of people capable of actually maintaining their agency are not common – recall he claims that “Of the prisoners only a few kept their full inner liberty.”¹⁴⁴ Interestingly, it seems that there is, in part, a biological explanation for this. Drawing on recent work by Baumeister and Gailliot, I argue that our ability to exercise self-control, as expressed in the previous chapter, is effected directly by the fact that we are embodied, and thus, because of our biological limitations. On Baumeister and Gailliot’s view:

¹⁴³ Perhaps it might be argued here that it would be unfair to say that these philosophers underestimate the power of these forces. It might, be argued that they acknowledge their power, but do not attach much significance to them. In either case, however, there is room in this respect for supplementation and deeper exploration in their accounts of human agency, which I explore in this chapter.

¹⁴⁴ V. Frankl, *Man’s Search for Meaning*, p. 76.

Self-control (or self-regulation) is the ability to control or override one's thoughts, emotions, urges and behaviour. Self-control allows for the flexibility necessary for successful goal attainment [...].¹⁴⁵

Furthermore, on Baumeister's "energy model," self-control "refers [...] to conscious efforts to alter [one's own] behaviour, especially restraining impulses and resisting temptations."¹⁴⁶ Baumeister and Gailliot's model thus maps on directly to the kind of self-control central to the self-control model of agency I defended in the first chapter. Importantly, on Baumeister and Gailliot's picture "self-control relies on glucose as a limited energy source" and "requires a certain amount of glucose to operate unimpaired."¹⁴⁷ They argue, I think convincingly, that "[t]he human body is undeniably an energy system, and its very life depends on ingesting energy and then using it to fuel activities, including complex psychological abilities,"¹⁴⁸ which include, specifically, self-control. Their research shows that the correlations between the body's glucose levels and the individual's ability to exercise self-control are dependent on one another in fundamental ways, showing that "decrements in self-control are caused in part by low glucose," and that "restoring glucose to higher and optimal levels should replenish the ability to exert self-control."¹⁴⁹

What is important about these studies is that they show how our ability to exercise the self-control associated with agency is directly linked to our embodiment –specifically to our blood glucose levels. What this means is that for the ordinary person, their efforts to exercise agency rely on an energy source that is depleted in the process of exercising self-control, thus making subsequent efforts at self-control more difficult to perform. For the ordinary person, access to sufficient food at appropriate times may be enough to broadly replenish energy stores which are necessary to exercise self-control, but even then, this may not always be the case. Baumeister and Gailliot give the following example: "Most obviously, dieting essentially involves restricting one's caloric intake, and there may be an ironic conflict in which the dietary restriction produces lower

¹⁴⁵ M. Gailliot, R. Baumesiter, N. DeWall, J. Maner, E. Plant, D., Tice, L. Brewer, and B. Schmeichel, 'Self-control Relies on Glucose as a Limited Energy Source: Willpower Is More Than a Metaphor', *Journal of Personality and Social Psychology*, vol. 92. no 2., 2007, p. 325.

¹⁴⁶ R. Baumeister as quoted by A. Mele, *Backsliding*, p. 105.

¹⁴⁷ M. Gailliot, et al., 'Self-Control Relies on Glucose as a Limited Energy Source' p. 325.

¹⁴⁸ *Ibid.*

¹⁴⁹ *Ibid.*, p. 326.

glucose, which in turn, undermines the willpower needed to refrain from eating.”¹⁵⁰ Those of us who have worked long hours without breaking to refuel, or have been on a diet, know how difficult it can become to resist temptations or control our urges and impulses while under these conditions.

Our bodies are vulnerable, and our agency is vulnerable because we are embodied. In this section I have shown that our embodiment thus poses a threat to our ability to exercise the kind of self-control associated with agency on the self-control view. Similarly, our embodiment makes human agency difficult to achieve. In this section I have shown only, however, that vulnerability and difficulty must at least be contingently necessary requirements for human agency. Later in this chapter I will argue for the deeper constitutive claims that vulnerability is constitutive of human agency, and that human agency should thus be thought of as difficult to achieve.

2.3 Self-Control and Situation

In the previous section I spoke about agency as vulnerable because the body is a situation in its own right. Being embodied is what enables action, and by extension agency, but our being embodied also renders our agency vulnerable. In this section, I focus on the following interrelated claim made by existentialists such as Beauvoir: the body is also always in a situation. As Beauvoir puts it, we are “part of the world of which [we are] a consciousness.”¹⁵¹ We not only live as embodied agents, using our bodies to express our intentions in action in the world, but our bodies are also engaged in a relationship with the world. We can retreat to the inner realms of our consciousness, but we cannot fully escape or transcend our material existence in this world. Recall Frankl’s suggestion that the concentration camp inmate was able to escape the external constraints of his situation by maintaining his inner liberty, that is, control over his internal mental and psychological states. For Beauvoir, the possibility of this retreat from the world is never absolute precisely because we nevertheless continue to be in the

¹⁵⁰ *Ibid.*, p. 334.

¹⁵¹ S.. Beauvoir, *The Ethics of Ambiguity*, Frechtman, B., trans., New York, Citadel Press, 1948, p. 7.

world – we always exist as bodies “crushed by the dark weight of other things.”¹⁵²

Similarly, perhaps one of the most important insights garnered from research in social psychology is that we are profoundly influenced by the situations in which we find ourselves. Though initially the point seems obvious, or perhaps even trivial, further research in social psychology has shown that, in practice, we find it difficult to estimate the influence of situation on an individual’s behaviour and actions. When we are asked to explain the actions of other people, we typically underestimate the role that situation plays in informing their actions. We also typically overestimate the role that character or personality plays. So pervasive are these findings in social psychology that this cognitive bias is now referred to as the “fundamental attribution error”¹⁵³ or the “correspondence bias.”¹⁵⁴ Philosophers, it seems, are not immune to this cognitive bias. For this reason, the extent of the impact that situations in which we find ourselves have on our ability to exercise agency is, I think, underexplored on accounts of agency such as Frankfurt and Watson’s (and in the philosophical discussions of agency following their work).

Work in experimental psychology, particularly in the experiments conducted by Milgram¹⁵⁵ and Zimbardo,¹⁵⁶ provide us with insight into the effects of situation on our ability to exercise the kind of self-control required for agency on the self-control view. Their experiments provide us with evidence about how people come to act against their own better judgements or intentions without even realising that their agency has been diminished by their situation. Importantly, these experiments provide us with further examples of situations in which our critical reflection is tainted by the influence of external forces. In this section, then, I argue that our agency is vulnerable because our ability to exercise

¹⁵² *Ibid.*

¹⁵³ L. Ross, and R. Nisbett, *The Person and the Situation: Perspectives of Social Psychology*, New York, McGraw-Hill, 1991.

¹⁵⁴ D. T. Gilbert, and E. E. Jones, ‘Perceiver-Induced Constraint: Interpretations of Self-Generated Reality’, *Journal of Personality and Social Psychology*, vol. 50, 1986, pp. 269–280.

¹⁵⁵ S. Milgram, ‘Behavioural Study of Obedience’, *The Journal of Abnormal and Social Psychology*, vol. 67, no. 4, 1963, pp. 371-378.

¹⁵⁶ C. Haney, C. Banks, and P. Zimbardo, ‘A Study of Prisoners and Guards in a Simulated Prison’, *Naval Research Reviews*, 9, Washington, Office of Naval Research, 1973, pp. 1-17.

self-control can be jeopardised by the situations in which we find ourselves – typically even without our own realisation.

In the Stanford Prison Experiment, Zimbardo and his research team created “a prison-like situation in which the guards and inmates were initially comparable and characterized as being ‘normal-average’.”¹⁵⁷ The aim of the simulation was to “observe the patterns of behaviour which resulted, as well as the cognitive, emotional and attitudinal reactions which emerged,”¹⁵⁸ in psychologically comparable persons in the scenario of imprisonment. Participants were paid a sum of money in return for work in the experiment in the role of either a guard or prisoner. In the simulated environment of a prison, prisoners were kept in barred cells and given activities similar to those given to prisoners in real prisons, in an attempt to give prisoners an experience which would produce “qualitatively similar psychological reactions”¹⁵⁹ to imprisonment which genuine imprisonment would elicit. The participants randomly assigned to be prisoners were told that they would have many of their civil rights (excluding only physical abuse) removed. However, “no other information” was given “about what to expect nor instructions about behaviour appropriate for a prisoner role.”¹⁶⁰ Similarly, while the guards were given the task of maintaining a “reasonable degree of order within the prison necessary for its effective functioning,”¹⁶¹ they were not given any further details about how this task was to be carried out; “they were intentionally given only minimal guidelines for what it meant to be a guard,” and were told only that there was a “categorical prohibition against the use of physical punishment or physical aggression.”¹⁶²

The findings of the experiment are some of the most influential findings from research in social psychology. Reflecting on the Stanford Prison Experiment 25 years after the study, Zimbardo summarises his findings as follows:

¹⁵⁷ *Ibid*, p. 3.

¹⁵⁸ *Ibid*.

¹⁵⁹ *Ibid*, p. 4.

¹⁶⁰ *Ibid*, pp. 6-7.

¹⁶¹ *Ibid*, p. 7.

¹⁶² *Ibid*.

The outcome of our study was shocking and unexpected to us, our professional colleagues, and the general public. Otherwise emotionally strong college students who were randomly assigned to be mock prisoners suffered acute psychological trauma and breakdowns. Some of the students begged to be released from the intense pains of less than a week of merely simulated imprisonment, whereas others adapted by becoming blindly obedient to the unjust authority of the guards. The guards, too – who also had been carefully chosen on the basis of their normal-average scores on a variety of personality measures – quickly internalized their randomly assigned role. Many of these seemingly gentle and caring young men, some of whom had described themselves as pacifists or Vietnam War “doves,” soon began mistreating their peers and were indifferent to the obvious suffering that their actions produced. Several of them devised sadistically inventive ways to harass and degrade the prisoners, and none of the less actively cruel mock-guards ever intervened or complained about the abuses they witnessed. Most of the worst prisoner treatment came on the night shifts and other occasions when the guards thought they could avoid the surveillance and interference of the research team.¹⁶³

From the observed behaviour of the guards we could draw moral insight about human nature. We might, for instance, remark on the ease with which the guards fell into aggressive and hostile behaviour. Zimbardo reports on the guards’ behaviour as follows:

Despite the fact that guards and prisoners were essentially free to engage in any form of interaction (positive or negative, supportive or affrontive, etc.), the characteristic nature of their encounters tended to be negative, hostile affrontive and dehumanizing [...]. Although it was clear to all subjects that the experimenters would not permit physical violence to take place, varieties of less direct aggressive behavior were observed frequently (especially on the part of the guards). In lieu of physical violence, verbal affronts were used as one of the most frequent forms of interpersonal contact between guards and prisoners.¹⁶⁴

I will not remark on the moral insights we could draw here about human nature, though there are certainly interesting parallels to be drawn with Hanna Arendt’s notion of ‘the banality of evil,’¹⁶⁵ on which Zimbardo himself draws in later

¹⁶³ P. Zimbardo, and G. Haney, ‘The Past and Future of U.S. Prison Policy Twenty-Five Years After the Stanford Prison Experiment’, *American Psychologist*, vol. 53, no. 7, 1998, p. 709.

¹⁶⁴ Haney, C., Banks, C., and Zimbardo, P., ‘A Study of Prisoners and Guards in a Simulated Prison’, p. 10.

¹⁶⁵ H. Arendt, *Eichmann in Jerusalem: a Report on the Banality of Evil*, London, Faber and Faber, 1963.

works such as *The Lucifer Effect*.¹⁶⁶ Here I would like to draw attention to the insight that situational influences on human behaviour can cause 'normal-average' personality types to behave in uncharacteristic ways. That is, the situation in which we find ourselves can influence not only our ability to act in accordance with our reflective intentions or better judgement, but the formation of those intentions and critical reflection itself. For the guards, their uncharacteristic behaviour was not explicitly required of the situation – for instance, the observed behaviour changes were not required in order to meet the objective of an effectively run prison, nor required for personal safety – in which case we could easily make sense of the otherwise 'normal-average' personality type's' behaviour in the situation. From the experiment, however, we see a change in both the guards and prisoners' characteristic behaviour due to situational factors which seemed to operate largely beyond their control. As Zimbardo puts it, "[t]he extremely pathological reactions which emerged in both groups of subjects testify to the power of the social forces operating."¹⁶⁷ So powerful were the forces operating that both guards and prisoners thought of life outside of the experimental situation as suspended. Participants began to assume their roles in private spaces, seemingly internalising the effects of their situation well beyond what would be expected of them in a research experiment, thus reflecting more than mere "public conformity or good acting."¹⁶⁸ Zimbardo explains the results as follows:

When the private conversations of the prisoners were monitored, we learned that almost all (a full 90 per cent) of what they talked about was directly related to immediate prison conditions, that is, food, privileges, punishment, guard harassment, etc. Only one-tenth of the time did their conversations deal with their life outside the prison [...]. The excessive concentration on the vicissitudes of their current situation helped to make the prison experience more oppressive for the prisoners because, instead of escaping from it when they had a chance to do so in the privacy of their own cells, the prisoners continued to allow it to dominate their thoughts and social relations. The guards too, rarely exchanged personal information during their

¹⁶⁶ P. Zimbardo, *The Lucifer Effect: Understanding How Good People Turn Evil*, New York, Random House, 2007.

¹⁶⁷ Haney, C., Banks, C., and Zimbardo, P., 'A Study of Prisoners and Guards in a Simulated Prison', p. 10.

¹⁶⁸ *Ibid*, p. 11.

relaxation breaks. They either talked about “problem prisoners,” other prison topics, or did not talk at all.¹⁶⁹

Zimbardo explains the results further, as follows:

The most dramatic evidence of the impact of this situation upon the participants was seen in the gross reactions of five prisoners who had to be released because of extreme emotional depression, crying, rage and acute anxiety. The pattern of symptoms was quite similar in four of the subjects and began as early as the second day of imprisonment [...] only two said that they were not willing to forfeit the money they had earned in return for being “paroled.” When the experiment was terminated prematurely after only six days, all the remaining prisoners were delighted by their unexpected good fortune. In contrast, most of the guards seemed to be distressed by the decision to stop the experiment and it appeared to us that they had become sufficiently involved in their roles that they now enjoyed the extreme control and power which they exercised and were reluctant to give it up [...]. None of the guards ever failed to come to work on time for their shift, indeed, on several occasions guards remained on duty voluntarily and uncomplainingly for extra hours – without additional pay.¹⁷⁰

Recall here my previous chapter’s discussion of Susan Wolf’s character JoJo, the evil dictator’s son. A large part of what I think is doing the work in her example of JoJo is the fact that the corrupting force which he is subject to what appears to be an explicitly corrupting force: the influence of his “heritage and upbringing”¹⁷¹ on his ability to exercise critical reflection and form his better judgements and intentions without being influenced, is significantly impeded by his father’s brainwashing. Wolf claims that what goes wrong in JoJo’s case is that he is not “connected to the world in a certain way,” which she explains only as “a desire that one’s self be controlled by the world in certain ways and not others.”¹⁷² It seems, however, that precisely which ways are understood as corrupting in this way and which ways are not, is still up for debate. The evidence from experiments in social psychology suggest some of the less intuitive and subtle ways in which a person’s agency, like JoJo’s, is threatened by situational factors. Some of the major influences on subjects in the Stanford Prison Experiment are

¹⁶⁹ *Ibid*, pp. 11-12.

¹⁷⁰ *Ibid*, p. 10.

¹⁷¹ S. Wolf, ‘Sanity and the Metaphysics of Responsibility’, p. 368.

¹⁷² *Ibid*.

typically, but I think wrongly, thought of as less corrupting of our agency because they are less obviously corrupting.

In the Stanford Prison Experiment, one of the major influencing factors on the prisoners and guards behaviour during the experiment was their attire. The research reveals that a seemingly low risk situational factor, such as how one is dressed, can have a dramatic impact on behaviour. In the experiment, both guards and prisoners were assigned uniforms: the guards' uniform consisted of "plain khaki shirts and trousers, a whistle, a police night stick (wooden batons), and reflecting sunglasses which made eye contact impossible," while the prisoners' uniform consisted of "a loose fitting muslin smock with an identification number on front and back, no underclothes, a light chain and lock around one ankle."¹⁷³ Interestingly, concentration camp uniforms outfitted by Nazis in the Second World War functioned similarly as identity erasers, and the uniform of domestic workers donned in Apartheid South Africa constituted those workers as subservient. Remarking on the influence this uniform was intended to have on participants, Zimbardo says:

The outfitting of both prisoners and guards in this manner served to enhance group identify and reduce individual uniqueness within the two groups. The khaki uniforms were intended to convey a military attitude, while the whistle and night-stick were carried as symbols of control and power. The prisoners' uniforms were designed not only to deindividuate the prisoners but to be humiliating and serve as symbols of their dependence and subservience.¹⁷⁴

And this is precisely the effect that the uniforms had on the subjects in the experiment. Remarking on the effect of the uniform on the prisoners, Zimbardo reports that the "ill-fitting uniforms made the prisoners feel awkward in their movements; since these 'dresses' were worn without undergarments, the uniforms forced them to assume unfamiliar postures, more like those of a woman than a man – another part of the emasculating process of becoming a prisoner."¹⁷⁵ Consequently, "prisoners adopted a generally passive response

¹⁷³ A Haney, C., Banks, and P. Zimbardo, 'A Study of Prisoners and Guards in a Simulated Prison', p. 8.

¹⁷⁴ *Ibid*, p. 8.

¹⁷⁵ *Ibid*, p. 8.

mode” culminating in “extreme emotional depression” and “acute anxiety.”¹⁷⁶ Zimbardo further remarks on the effect of the uniform on the guards by claiming that the guards immediately adopted a “very active initiative role in all interactions” typically characterised by domination and aggression. Importantly, Zimbardo says:

[A] video-taped analysis of total guard aggression showed a daily escalation even after most prisoners had ceased resisting and prisoner deterioration had become obvious to them. Thus, guard aggression was no longer elicited as it was initially in response to perceived threats, *but was emitted simply as a “natural” consequence of being in the uniform of a “guard” and asserting the power inherent in that role* [...] [A]nother guard who detained an “incorrigible” prisoner in solitary confinement beyond the duration set by the guards’ own rules [...] conspired to keep him in the hole all night while attempting to conceal this information from the experimenters who were thought to be too soft on the prisoners.¹⁷⁷

Similar results with regards to these intuitively less explicitly corrupting forces are at play in the results from the famous Milgram Experiment. In the Milgram Experiment, which was designed to test a subject’s obedience to authority, subjects were deceived into thinking that they had been recruited by Yale University to participate in a study of the “effects of punishment on memory.” Under the gaze of an actor whose role assumed that of a ‘scientist,’ the subject assigned the role of ‘teacher’ was instructed to administer electric shocks to the ‘learner’ when he incorrectly answered questions. The ‘learner,’ who was really an actor, and not receiving any actual electric shocks, was introduced to the ‘teacher’ as just another recruit for the Yale research on punishment and learning. The ‘teacher’ was then falsely led to believe that their positions at teachers or learners had been randomly assigned. The ‘teacher’ subject was placed into a separate room with the ‘scientist’ and two machines: the machine with which the subject was to communicate with the victim, who was in another room, and a machine used to administer the electric shocks, both of which the subject was expected to operate. The machine to administer the electric shocks had clearly marked voltage levels (ranging from 15 to 450 volts) and verbal

¹⁷⁶ *Ibid*, p. 10.

¹⁷⁷ *Ibid*, p. 12, my emphasis.

designations ranging from “Slight Shock to Danger: Severe Shock.” The ‘teacher’ was expected to increase the voltage administered to the ‘learner’ for incorrect answers. As the experiment progressed, the ‘teacher’ was expected to administer shocks reaching the most severely painful voltage. The verbal reactions of the ‘learner’ to the electric shock could be heard by the teacher, and increased in volume and intensity as the voltage of the shock was increased (though in reality the actor playing the learner was merely pretending to scream in increasing pain). Milgram explains the interesting position the ‘teacher’ subjects found themselves in, in terms of an internal conflict, as follows:

The subject is placed in a position in which he must respond to the competing demands of two persons: the experimenter and the victim. The conflict must be resolved by meeting the demands of one or the other; satisfaction of the victim and the experimenter are mutually exclusive. Moreover, the resolution must take the form of a highly visible action, that of continuing to shock the victim or breaking off the experiment.¹⁷⁸

Reporting on the findings from his experiment, Milgram says that “[o]ne might suppose that a subject would simply break off or continue as his conscience dictated. Yet this is very far from what happened.”¹⁷⁹ Milgram explains what the results showed, as follows:

the sheer strength of obedient tendencies manifested in this situation. Subjects have learned from childhood that it is a fundamental breach moral conduct to hurt another person against his will. Yet, 26 subjects abandon this tenet in following the instructions of an authority who has no special powers to enforce his commands. To disobey would bring no material loss to the subject; no punishment would ensue. It is clear from the remarks and the outward behaviour of many participants that in punishing the victim they are often acting against their own values, subjects often expressed deep disapproval of shocking a man in the face of his objections, and others denounced it as stupid and senseless. Yet the majority complied with the experimental commands.¹⁸⁰

The results from the Milgram Experiment, similar to those of the Stanford Prison Experiment, show the tremendous impact of situational factors, which are often

¹⁷⁸ S. Milgram, ‘Behavioural Study of Obedience’, p. 378.

¹⁷⁹ *Ibid*, p. 377.

¹⁸⁰ *Ibid*, p. 376.

thought of as less explicitly corrupting, on behaviour – the way people around us are dressed and the influence of authority. Although there has been some criticism of the conclusions of these experiments, the basic conclusions discussed here remain relatively uncontroversial.¹⁸¹ From my discussion in this section I think that we can conclude that the corruptive potential of our situation is gradated, and while those forces which seemingly operate at the lower end of the scale are typically taken to be less corruptive, they are nevertheless significant factors which influence our behaviour and threaten to undermine our agency in ways which are equally effective and deleterious as those operating at the upper end of the scale. What the Stanford Prison Experiment and the Milgram Experiment partly show us is that even the seemingly smallest modifications in physical and social settings can have powerful effects on our behaviour. Furthermore, these seemingly less corrupting forces are far more difficult for us to identify, and thus to overcome, in everyday life. The evidence from experiments in social psychology suggests that there are implicitly far more every-day, routine, and commonly experienced ways in which our situation influences the formation of our better judgements and our critical reflection.

Self-control, it would seem, is more difficult for us to achieve than is commonly suggested on analytic philosophical accounts such as Frankfurt and Watson's, given the pervasive and corrupting potential of circumstances beyond our control. It is our situation in the world, I have suggested in this section, which makes our agency both vulnerable and difficult to achieve. Our situation in the world poses a threat to our ability to exercise the kind of self-control necessary for human agency on the self-control view. In this section I have shown only, however, that vulnerability and difficulty must at least be contingently necessary requirements for human agency. Later in this chapter I will argue for the deeper constitutive claims that vulnerability is constitutive of human agency, and that human agency should thus be thought of as difficult to achieve.

¹⁸¹ Even if there are some important questions to be raised about the validity of these studies, the less controversial conclusions which I discuss make both experiments illustrative and informative for my philosophical investigation into the nature of human agency.

2.4 Cognitive Limitations for Self-Control

In the related fields of cognitive science and behavioural economics, recent research has increasingly exposed a number of flawed heuristics and cognitive biases which affect a person's decision making ability – or their ability to form their better judgments or intentions. Research in these fields has shown that there are predictable flaws in our judgements, which lead to actions caused by an identifiable “mental error,” namely, the employment of a flawed heuristic mechanism resulting in cognitive bias.¹⁸² In this section I will argue that because human agents employ flawed heuristics and have cognitive biases, their ability to exercise self-control is at risk and thus human agents are to be seen as vulnerable in light of this cognitive vulnerability.

In their now seminal paper “Judgment under Uncertainty: Heuristics and Biases,” Tversky and Kahneman, who could be considered the fathers of behavioural economics, explain the role of flawed heuristics and biases in judgement and decision making, as follows:

A substantial body of research in cognitive psychology and decision making is based on the premise that these cognitive limitations cause people to employ various simplifying strategies or rules of thumb (heuristics) to ease the burden of mentally processing information to make judgments and decisions. These simple rules of thumb are often useful in helping us deal with complexity and ambiguity. Under many circumstances, however, they lead to predictably faulty judgments known as cognitive biases.¹⁸³

In the previous section, I argued that part of what makes human agency a philosophically interesting topic is the fact that, as human beings, we have very complex mental lives. This complexity is important because it gives rise the mental conditions that are necessary for us to exercise the kind of self-control central to human agency. But, this complexity is also the possible source of the inability to exercise self-control because our brains employ various mental shortcuts to deal with the level of complex mental processing required. These

¹⁸² A. Tversky, and D. Kahnemann, ‘Judgment under Uncertainty: Heuristics and Biases’, *Science, New Series*, vol. 185, no 4157, 1974, pp. 1124-1131.

¹⁸³ *Ibid*, pp. 1124.

mental shortcuts are manifest in our daily lives in the form of cognitive biases, which inform our critical reflection and evaluative judgements in ways that are not under our conscious control. Tversky and Kahneman explain cognitive biases as follows:

Cognitive biases are mental errors caused by our simplified information processing strategies. It is important to distinguish cognitive biases from other forms of bias, such as cultural bias, organizational bias, or bias that results from one's own self-interest. In other words, a cognitive bias does not result from any emotional or intellectual predisposition toward a certain judgment, but rather from subconscious mental procedures for processing information. A cognitive bias is a mental error that is consistent and predictable.¹⁸⁴

The well known example they give of a cognitive bias is our ability to make judgements about the apparent distance of objects, which they explain as follows:

The apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. This rule has some validity, because in any given scene the more distant objects are seen less sharply than nearer objects. However, the reliance on this rule leads to systematic errors in estimation of distance. Specifically, distances are often overestimated when visibility is poor because the contours of objects are blurred. On the other hand, distances are often underestimated when visibility is good because the objects are seen sharply. Thus the reliance on clarity as an indication of distance leads to common biases [...]. This rule of thumb about judging distance is very useful. It usually works and helps us deal with the ambiguity and complexity of life around us. Under certain predictable circumstances, however, it will lead to biased judgment.¹⁸⁵

Drawing on this work in flawed heuristics and cognitive biases, behavioural economists, such as Dan Ariely, maintain that is our ability to exercise rational reflection, because our rationality is itself a "bounded rationality."¹⁸⁶ He argues that "while we have the capacity for rational choice and behaviour we are also

¹⁸⁴ *Ibid*, pp. 1124.

¹⁸⁵ *Ibid*, pp. 1124.

¹⁸⁶ The idea of bounded rationality here is used to contrast with the view of the essentially and infinitely rational agent assumed in the standard economic models.

deeply flawed,”¹⁸⁷ and that our errors in mental processing can be easily predicted. He explains:

The conclusions from behavioural economics suggest that the ways in which we are irrational are neither random, nor senseless. They are easily explained because they are also systematic, and since we repeat them again and again, we are also predictably irrational. We make the same types of mistakes over and over, because of the basic wiring of our brains [...]. Behavioural economists believe that people are susceptible to irrelevant influences from their immediate environment, irrelevant emotions, short-sightedness and other forms of irrationality.¹⁸⁸

In *Predictably Irrational*, Ariely explores the ways in which flawed heuristics and cognitive biases influence the decisions we make and the actions we take in our daily lives. Among other things, he argues that “what we think, often with great confidence, influences” our “decisions in daily life,”¹⁸⁹ but that it is quite often not what actually influenced those decisions. In daily life, he argues that when it comes to efforts at self-control – such as dieting, shopping and saving money – we are less in control than we typically take ourselves to be. For Ariely, we act on the basis of flawed heuristics or cognitive biases far more often than we like to think. Because these flawed heuristics and biases take place in our brains beyond our conscious rational control, the effect of a flawed heuristic or cognitive bias on our agency is deleterious.

For example, Ariely discusses the role of the “decoy effect,” and the “asymmetric dominance effect,” on daily decision making practices. He explains that we typically rely on comparisons when making assessments and choices – our choices are generally made in a context in which we can compare at least one option with another. Thus, our ability to form intentions or better judgements relies, to some extent, on our ability to effectively assess and choose between our options by comparing them with one another. Furthermore, according to Ariely, “we not only tend to compare things with one another but also tend to focus on comparing things that are easily comparable – and avoid comparing things that

¹⁸⁷ D. Ariely, *Predictably Irrational: The Hidden Forces that Shape our Decisions*, New York, Harper, 2008, p. xi.

¹⁸⁸ *Ibid.*

¹⁸⁹ *Ibid.*

cannot be compared easily.”¹⁹⁰ As Ariely puts it, “[w]e always compare jobs with jobs, vacations with vacations, lovers with lovers and wines with wines.”¹⁹¹ The heuristic mechanism, which helps us to sort through the many choices we have to make, breaks down when we are faced with choices with which we need to compare options that are not easily compared. It is the resultant cognitive bias from this flawed heuristic which makes us particularly vulnerable to what is called the decoy effect.

Ariely uses the example of shopping for a house and being shown three different, though equally priced, options by a real estate agent. Two of the houses are similar in features (style A), while one has unique features (contemporary style B), quite different from the first two. One of the A style houses (-A), however, is in need of a new roof.¹⁹² Because the comparison between A and B represents such different options, the choice between them is difficult, and, by comparison, the choice between A and -A is relatively simpler. According to Ariely’s research, because it is easier for us to compare the two A style houses, we predictably, and reliably, dismiss the B style house because we do not have another B style house with which to compare it. In the example, house -A represents what can be seen as the “decoy.” House -A thus “creates a simple relative comparison with (A) look better, not just relative to A, but overall as well.”¹⁹³ It is precisely our vulnerability to this decoy effect that marketing gurus and salesmen alike strive to exploit. As much is illustrated in Ariely’s example of Sam, a television salesman:

Take Sam the television salesman. He plays the same kind of trick on us when he decides which televisions to put together on display:

36-inch Panasonic for \$690

42-inch Toshiba for \$850

50-inch Philips for \$1,480

[...]. In this case, Sam knows that customers find it difficult to compute the value of different options. (Who really knows if the Panasonic at \$690 is a better deal than the Philips at \$1,480?) But Sam also knows that given three choices, most people will take the middle choice [...]. So guess which television Sam prices as the middle option? That’s right – the one he wants to sell!¹⁹⁴

¹⁹⁰ *Ibid*, p. 8.

¹⁹¹ *Ibid*, p. 7.

¹⁹² *Ibid*, p. 8.

¹⁹³ *Ibid*, p. 9.

¹⁹⁴ *Ibid*, pp. 3-4.

Ariely's sampling is but a drop in the ocean of the literature on flawed heuristics and biases that has developed over the last half century. While he discusses a host of other cognitive biases and flawed heuristics, which influence our choices and behaviours, what is crucial for my discussion here is that the mental processes resulting in our choices or actions occur beyond our control – unconsciously and systematically. It is true of course that not all of our cognitive biases will necessarily result in us making *unwanted* judgements. For example, even if I was biased by Sam's decoy, it might be the case that the Toshiba fits the exact requirements for the television set that I set out to buy in the first place – even if I have fallen prey to the decoy effect, I still choose an option I really want to choose, all things considered. Moreover, I might be the sort of person who enjoys allowing a salesperson to influence their decisions about what to purchase (salespeople may have our best interests at heart after all, might they not?).

However, there certainly are times when our cognitive biases lead us to “have an unwanted judgment, emotion, or behaviour because of mental processing that is unconscious or uncontrollable.” That is to say, “the person making the judgment [would] prefer not to be influenced in the way he or she was.”¹⁹⁵ Wilson and Brekke, rather aptly I think, have termed this phenomenon “mental contamination.”¹⁹⁶ They claim that the analogy of contamination is useful for discussing cognitive biases because:

it focuses on the difficulty of avoiding many biases. Something that is contaminated is not easily made pure again, which we believe is an apt metaphor for many mental biases. We argue that, because of a lack of awareness of mental processes, the limitations of mental control, and the difficulty of detecting bias, it is often very difficult to avoid or undo mental contamination.¹⁹⁷

Unlike physical contamination, which, for the most part, is easily detected through the observable symptoms in the body such as a runny nose when sick

¹⁹⁵ T. Wilson, and N. Brekke, 'Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations', *Psychological Bulletin*, vol. 116, no. 1., 1994, p. 117.

¹⁹⁶ *Ibid.*

¹⁹⁷ *Ibid.*

with the flu, mental contamination rarely results in such easily observable symptoms, and this makes mental contamination difficult for us to detect, and thus difficult to overcome. According to Wilson and Brekke:

The same is often true, of course, in the physical realm: People cannot directly observe the processes that cause physical contamination, such as the invasion of their cells by rhinoviruses. In the physical realm, however, they are often observable symptoms of contamination, even when the process of contamination is unobservable. Although people cannot observe rhinoviruses, a stuffed-up nose tells them they have a cold. If one is wondering whether a gallon of milk is fresh or spoiled, a quick whiff will reveal the answer. There are seldom such observable symptoms, such as smell, temperature, or physical appearance, indicating that a human judgment is 'spoiled' [...].

Sometimes, we acknowledge, there are clues that a judgment has been contaminated. This is most likely when people have an unexpected reaction to a stimulus. If Jim enjoys a movie much less than he expected, he might be suspicious that he was influenced by the people behind him slurping their drinks [...]. These clues are not definitive, however, and can be misleading. Jim has no way of knowing how he would have felt about the movie in the absence of the slurping noises [...]. Furthermore, it is likely that many cases of mental contamination do not have such large effects that people have very unexpected reactions. It is probably more common for the effects to be subtle (e.g., giving a favourite student a B instead of a C). In these cases, the bias is likely to go completely unnoticed, without a moment's doubt about the validity of the judgment.¹⁹⁸

The problem of detection is further compounded by the fact that the mental processing, which results in an error of judgement, largely occurs unobserved by the person making the judgement – that is, unconsciously. Unconscious sources of judgement such as flawed heuristics and biases are therefore difficult for us to detect in and for ourselves. As Wilson and Brekke explain further:

One difference between mental and physical contamination, however, concerns their ease of detection. It is much more difficult to detect mental contamination, particularly at the individual level.

One impediment to such recognition is people's limited access to their mental processes [...]. Despite the stormy history of the literature on unconscious processing, the idea that people are unaware of a substantial amount of their mental processing has a firmer toehold in social and cognitive psychology than ever before. Because of this limited access, mental processes leading to contamination can occur unobserved.¹⁹⁹

¹⁹⁸ *Ibid.*, p. 121.

¹⁹⁹ *Ibid.*

The detection problem emphasises the difficulty of eliminating the effects of unconscious mental processing on our choices and actions because our lack of awareness of cognitive biases means that we do not take any of the necessary steps to debias ourselves (if debiasing is possible). As Wilson and Brekke argue, “[t]he fact that mental contamination is difficult to detect makes it hard to avoid or eliminate, for the simple reason that if people are unaware that their judgment is biased, they will not try to debias it.”²⁰⁰

Interestingly, Marilyn Friedman suggests, I think convincingly, that others around us might be better positioned to point out the effects of our biases on our choices and actions.²⁰¹ With someone else’s assistance, we might then sidestep, to some extent, the problem of detection, and thus improve our chances of overcoming bias: biases identified by third parties may be easier to overcome, precisely because they are easier to detect when we have a little help from our friends. Friedman explains as follows:

For good psychological reasons, each person’s unaided thinking cannot be trusted to discern its own biases. One’s own thinking – explicit and implicit, avowed and tacit – is not fully transparent to oneself. Biases are recognisable, for example, in the particular moral problems which attract or which escape one’s attention, by the metaphors chosen to express oneself, and by one’s reactions to different sorts of persons. The beneficiaries, victims, observers, and so on, of one’s behaviour may be better situated than oneself to discern biases hidden behind one’s articulated moral attitudes, because they can comprehend those avowals contextually in the light of one’s related actions and practices.²⁰²

However, even with the possibility of help from some third party in identifying our biases, cognitive biases are notoriously difficult to overcome. Even when we have identified a particular bias, it will sometimes be impossible for us not to be

²⁰⁰ *Ibid*, p. 122.

²⁰¹ Friedman is not talking explicitly about cognitive biases understood as the result of a flawed heuristic, rather she is discussing what we might call moral biases or prejudices. Nevertheless, her suggestion seems plausibly transferred across the discourses. Indeed, in some of Lowenstein’s later work on debiasing techniques investigate this suggestion to some extent. I will return to this idea later. M. Friedman, ‘The Impracticality of Impartiality’, *The Journal of Philosophy*, vol. 86, no. 11, 1989, pp. 645-656. See also, L. Babcock, G. Loewenstein, and S. Issacharoff, ‘Creating Convergence: Debiasing Biased Litigants’, *Law & Social Inquiry*, vol. 22, no. 4, 1997, pp. 913-925.

²⁰² M. Friedman ‘The Impracticality of Impartiality’, p. 656.

unduly influenced in our decisions and actions by certain cognitive biases. Analogously, while we might be fully aware of an optical illusion, we cannot overcome its effects. Tversky and Kahneman explain:

Cognitive biases are similar to optical illusions in that the error remains compelling even when one is fully aware of its nature. Awareness of the bias, by itself, does not produce a more accurate perception. Cognitive biases, therefore, are, exceedingly difficult to overcome.²⁰³

Similarly, Wilson and Brekke argue that the effects of mental contamination on behaviour are so efficacious that they are notoriously difficult to overcome. They explain, as follows:

Even if people are aware that an unwanted mental process is tainting their judgments and recognize the direction and magnitude of the resulting bias, they need to be able to control their responses sufficiently to correct the bias. Unfortunately, however, a considerable amount of recent research suggests that people's ability to control their thoughts and feelings is limited. Although people can stop certain kinds of mental operations from occurring, such as simple arithmetic operations, many other kinds of mental processes are extremely difficult to control once they are set in motion, especially those that occur outside awareness. Thus, avoiding mental contamination by stopping a sequence of thoughts or mental operations is unlikely to be a consistently successful strategy.

Once a mental process has run its course, resulting in a belief (e.g., 'Hernandez's paper is of C quality') or feeling (e.g., sadness), it is very difficult to erase that belief or feeling. Beliefs and feelings change as people gain new information, of course. Reading a glowing letter of recommendation can change one's mind about a job applicant, and sadness can change to euphoria on the receipt of good news. It is very difficult, however, to erase a belief or feeling, in the absence of new information, by sheer will.²⁰⁴

Wilson and Brekke's argument, I think, serves to highlight the ways in which our cognitive vulnerability contributes to the difficulty of achieving the kind of self-control associated with agency on the self-control view. In light of these cognitive concerns, our ability, as human agents, to exercise self-control must thus be seen as vulnerable. In this section I have shown only, however, that vulnerability and difficulty must at least be contingently necessary requirements for human

²⁰³ A. Tversky, and D. Kahneman, 'Judgment under Uncertainty', pp. 1124.

²⁰⁴ T. Wilson, and N. Brekke, 'Mental Contamination and Mental Correction', p. 122.

agency. Later in this chapter I will argue for the deeper constitutive claims that vulnerability is constitutive of human agency, and that human agency should thus be thought of as difficult to achieve.

2.5 Self-Control and Internal Conflict

In the previous chapter, I argued that a human agent is someone who experiences conflict between motivations and intentions. I argued that self-control is not simply a structural feature of human agency, but rather the central psychological activity in which we must actively engage on an ongoing basis in order to be agents. In addition, I argued that this psychological activity requires that we experience conflict and constraint. On the one hand, this internal complexity allows for human agency. On the other hand, as I will argue in this section, the presence of internal conflict makes the self-control required for human agency difficult to achieve. Such difficulty is a consequence of two distinct reasons, both of which allow me to argue that the very conditions which give rise to human agency in terms of internal conflict, make our agency itself vulnerable. In this section I argue that because internal conflict is central to human psychology, the self-control necessary for human agency is difficult to achieve and human agents are vulnerable. It is because of our internal division that we are subject to what I call internal vulnerability. Because human subjects experience internal conflict between motivations and intentions, our self-control is constantly at risk.

2.5.1 The Fight for Self-Control is not Fought on Even Ground

In the first chapter, I argued, following Midgley, that the kind of mental or psychological activity associated with self-control amounts to an “organisation of the inner crowd.”²⁰⁵ That is, we are able to manage the conflicts (for example, between and amongst our different desires and intentions²⁰⁶), which occur within our own mental complex. Midgley suggests, I think rightly, that it is in fact

²⁰⁵ M. Midgley, *The Ethical Primate*, p. 183.

²⁰⁶ I will return to this in Chapter 4 in more detail.

notoriously difficult to achieve effective management of the inner crowd. In part, this difficulty is due directly to the presence of internal constraints, which arise within our own psyche. There are a number of elements, which arise within our own psychological make-up, that provide examples of internal constraints to our ability to exercise self-control, and that threaten our ability to effectively manage the inner crowd. We have all been overwhelmed by some desire or motivation which prevents us from being able to exercise self-control. When we are able to exercise self-control, in the face of constraints of this type, we are aware of the immense effort involved. In research from behavioural economics, George Lowenstein has most influentially argued that the affect of what he calls “visceral factors” on our choices and actions bring into question the extent of control we have over those choices and actions. In Lowenstein’s view, then, when organising the inner crowd, we are often unable to control how much certain desires or inclinations motivate us, which thus limits our ability to exercise the self-control. According to Lowenstein, visceral influences undermine our self-control more often than we typically assume they do.

His research suggests that the effect of visceral factors on our behaviour can be ranked along a gradient according to intensity – ranging from having a limited effect on behaviour to a fully determinative effect on behaviour regardless of the person’s better judgements or intentions. According to Lowenstein:

People often act against their self-interest in full knowledge that they are doing so; they experience a feeling of being “out of control” [Due to] “visceral factors,” which include drive states such as hunger, thirst and sexual desire, moods and emotions, physical pain, and a craving for a drug one is addicted to. At sufficient levels of intensity, these and other visceral factors, cause people to behave contrary to their own long-term self-interest, often with full awareness that they are doing so.²⁰⁷

And again, he says:

The overriding of rational deliberation by the influence of visceral factors is well illustrated by the behavior of phobics who are typically perfectly aware that the object of their fear is objectively

²⁰⁷ G. Loewenstein, ‘Out of Control: Visceral Influences on Behavior’, *Organizational Behaviour and Human Decision Processes*, vol. 65, no. 3, 1996, pp. 272-273.

nonthreatening, but are prevented by their own fear from acting on this judgment.²⁰⁸

At the low extreme of this gradient, weak visceral factors have less impact on behaviour, and self-control can still be exercised. For example, “someone who is *slightly sleepy* might decide to leave work early or to forgo an evening’s planned entertainment so as to catch up on sleep.” However, this person could also just as easily take a stimulant and stay awake for the movie, planning to go to bed straight after the show. At the high extreme of this gradient, though, the effect of a visceral factor is so powerful that it can be seen to undermine a person’s ability to exercise the kind of self-control necessary for agency. Lowenstein says:

Finally, at even greater levels of intensity, visceral factors can be so powerful as to virtually preclude decision making. No one decides to fall asleep at the wheel, but many people do.²⁰⁹

Of course, we should expect visceral factors to influence our behaviour in a number of ways. As Lowenstein puts it, “[r]ational choice requires that visceral factors be taken into account. It makes good sense to eat when hungry, to have sex when amorous, and to take pain killers when in pain.”²¹⁰ He goes on to emphasise that, “[i]ncreases in the intensity of visceral factors, however, often produce clearly suboptimal patterns of behaviour,”²¹¹ such as those which go against our better judgement, but over which we are not able to exercise self-control. Because of the powerful effect of some visceral factors, our ability to manage our inner crowd is made difficult: some of the competing desires and motivations on which we critically reflect can overpower the judgements and intentions based on those reflections.

2.5.2 In Every Victory of Self-Control there Lies Defeat

Recall that in the first chapter I outlined a basic account of human internal conflict which can be drawn from Freud. On Freud’s account, the mind is seen as

²⁰⁸ *Ibid.*, p. 273.

²⁰⁹ *Ibid.*, p. 273.

²¹⁰ *Ibid.*, p. 273.

²¹¹ *Ibid.*, p. 273.

inherently motivationally divided; and following Freud I argued that a constitutive feature of human agency is the ability to respond to this internal conflict in order to act intentionally. Similarly, as sketched in section 2.1, Nietzsche argues that human agency involves responding to internal conflict in order to perform intentional action. According to Nietzsche, moreover, asserting our agency through self-control is often simultaneously a great victory and a frustrating defeat. When we are motivated by our better judgement, we succeed only in defeating a part of ourselves that we recognise as perhaps motivating us to do otherwise. For Nietzsche, the condition for “self-overcoming” necessary for agency is at once a “self-negating and yet self-identifying and self-affirming, state.”²¹² In order to explain Nietzsche’s idea further, I will return to what I noted at the beginning of this chapter. That is, Nietzsche agrees that the fundamental psychological steps involved in exercising the kind of self-control necessary for agency requires that, as agents, we form values on the basis of our basic responses to the world or brute desires. This process, in turn, for the agent, involves selecting and affirming or endorsing some desires over others with which they may be in conflict. However, Nietzsche notes that, if we are to genuinely affirm or endorse some desires over others, we must, at the very least, acknowledge the fact that the satisfaction of our endorsed desires precludes the fulfilment of other desires. If we do not acknowledge these conflicting desires as our own desires, which now cannot reach fulfilment, in a very important sense, we deceive ourselves about ourselves, and are thus guilty, as Nietzsche puts it, of *ressentiment*. For him, *ressentiment* is, at least in part, a psychological condition which results in a fractured and damaged individual, who lacks integrity and sovereignty. The “man of *ressentiment*” fails to understand that the genuine endorsement of a value requires the acknowledgment that in his fulfilment of a value, he has other desires which then remain unsatisfied, but that these desires are no less his own. By denying that these desires are his own, he deceives himself about himself. The man of *ressentiment* fails to acknowledge that certain of his own desires cannot be fulfilled, and in doing so, fails to understand the implications of the endorsement of a value on his life. He thus fails, according to Nietzsche, to genuinely endorse the value at all. Reginster explains, as follows:

²¹² R. Pippin, ‘How to Overcome Oneself’, p. 82.

There is no genuine endorsement of a value, therefore, without the acknowledgment of those of our desires which conflict with its realization. To acknowledge the presence of conflicting desires and to accept the fact that they have to be left unsatisfied demands unflinching honesty with ourselves. But the required honesty is precisely what the “man of *ressentiment*” lacks.²¹³

Not only, then, are we subject to internal struggles, but the resolution of these struggles, through the process of critical reflection and the formation of our better judgement, involves our coming to terms with the reality of this resolution. In all victories resulting in self-control, we manage simply to defeat another part of ourselves. On the one hand, of course, we might see this victory over ourselves as worthwhile – it gives us agency, control and freedom. On the other hand, there is some part of us that is defeated, denied and repressed. If we are not then sometimes guilty of *ressentiment*, which challenges our agency in fundamental ways,²¹⁴ then we are, in some sense, both liberated and frustrated by exercising self-control. We have all felt this frustration build and break in moments where self-control escapes us or when we act in bold defiance of our intentions (particularly when they align with our better judgement as in cases of orthodox self-control) with some secret satisfaction. To have control over all this places enormous pressure on any agent, and, again, shows the fragile internal situation we must navigate in order to exercise our agency.

In light of both 2.5.1 and 2.5.2, our agency can be seen as vulnerable when our internal constraints make the kind of self-control necessary for agency on the self-control view difficult to achieve. Furthermore, and I think interestingly, this difficulty and vulnerability is the direct result of the complexity of our mental lives, which provides the very conditions which make self-control possible in the first place. Drawing on this, in the next section I argue that we can offer some normative prescriptions for improving our agency in the face of this difficulty and vulnerability.

²¹³ B. Reginster, ‘Nietzsche on *Ressentiment* and Valuation’, *Philosophy and Phenomenological Research*, vol. LVII, no. 2, 1997, pp. 281-305.

²¹⁴ For a detailed discussion of this see my, ‘Questions of the Self in the Personal Autonomy Debate’.

2.6 Improved Self-Control

In this section, I will argue that not only is a robust account of the self-control view of agency achieved by making these additions, but the inclusion allows for the potential for us to develop strategies to improve our agency. If we can identify the ways in which our agency is vulnerable, as well as the factors which make our agency difficult to achieve, we can begin to formulate action-guiding agential principles. In this section, then, I turn my attention to some of the ways in which we might improve our self-control. In doing so, I will offer some basic action-guiding practices. My suggestions here are undoubtedly only a start, and there remains considerable room for research in the future on this topic.

We devote time and effort to finding ways to overcome our vulnerability because, to some extent, we are already aware, in our daily lives, that our ability to exercise the kind of self-control necessary for agency is vulnerable. As Alfred Mele puts it:

If the popularity of self-help manuals and therapy is any indication, irrational behaviour is remarkably common. People eat, spend, smoke, and so on much more (or less) than they think they should, and they spend much time and effort attempting to bring their behaviour into line with their better judgments.²¹⁵

According to Mele, there are a number of ways we can learn to exercise self-control more effectively, even in the face of the constraints which make our agency vulnerable. Mele points out the following:

In normal agents, a capacity for self-control is not a mental analogue of brute physical strength. We learn to resist temptation by promising ourselves rewards for doing so, by vividly imagining undesirable effects of reckless conduct, and in countless other ways. Our powers of self-control include a variety of skills—and considerable savvy about which skills to use in particular situations.²¹⁶

²¹⁵ A. Mele, *Irrationality*, p. vii.

²¹⁶ A. Mele, *Backsliding*, pp. 93-94.

Part of what makes us rational is the struggle against irrationality, so irrationality is ineradicable, although, in most cases but perhaps not in all, we should fight it when it appears. Similarly, there are a number of active steps people can take to improve their chances of successfully exercising the kind of self-control necessary for agency. Employing techniques to reduce the impact of factors such as bodily, situational, cognitive and internal vulnerability might help to reduce the extent to which our agency is vulnerable. Here, as with many things, understanding is the crucial first step.

Recall my discussion of Baumeister's research on the 'energy model' of self-control. Baumeister suggests that our ability to exercise the kind of self-control associated with agency is directly linked to our blood glucose levels. His research, as I have already indicated, suggests that a decrease in blood glucose levels makes self-control more difficult to achieve. However, further research has shown that we can mitigate the potentially deleterious effects of low blood glucose levels on our decision-making by, for example having a snack. In their research, Shai Danziger, Jonathan Levav and Liora Avnaim-Pesso, found that eating food, and thereby restoring blood glucose levels, increased parole officers' ability to make fair judgements in cases where they had to make numerous parole decisions in a day.²¹⁷ Much like Baumeister's research, Danziger et al. show that "self-control operates on the basis of a limited recourse, akin to energy or strength, that can become depleted through use," and that the "performance" of self-control "grow[s] worse during consecutive or continuous efforts, just as a muscle becomes tired."²¹⁸ Here, Baumeister's research has further implications for our ability to exercise self-control: his research shows that the ability to exercise self-control in the face of temptations or influences from external forces becomes stronger with increased use. Mele argues that the results from this research provide "encouraging news"²¹⁹ for those of us who want to learn to exercise self-control more effectively, because the results show that self-control improves with practice. In Baumeister's research, "[p]articipants who exercised

²¹⁷ S. Danziger, J. Levav, and L. Avnaim-Pesso, 'Extraneous Factors in Judicial Decisions', *Proceedings of the National Academy of Sciences of the United States of America*, Princeton, vol. 108, no. 17., 2011, pp 6889-6892.

²¹⁸ R. Baumeister as quoted in A. Mele, *Backsliding*, pp. 130-131.

²¹⁹ *Ibid*, p. 134.

self-control in various ways over a two-week period later displayed significantly more self-control than a control group.”²²⁰ By keeping our blood glucose levels at the right level, and importantly, by training our willpower as we would train a muscle for added strength and endurance, it seems that we can improve the success of our efforts at self-control, and so learn to exercise our agency more effectively.

As we know from our own experiences, and the experiences of others, we typically already do adopt some strategies for resisting temptations from external forces (such as situational forces discussed in section 2.3) and internal forces (such as the effect of visceral factors discussed in section 2.5). Drawing on Baumeister’s research, Mele argues that we are better at exercising self-control when the stakes are high. He claims that research shows that we implicitly rank the importance of successfully exercising self-control in any situation, and that we are likely to keep energy in reserve in order to assist our motivational strength when important issues of self-control arise.²²¹ But, Mele also argues that there are a number of active techniques we can learn and employ when resisting temptations, thereby improving our self-control.

Recall, from section 2.5, my discussion of Lowenstein’s research on the effect of visceral factors on decision making. According to Lowenstein, the effect of strong visceral factors resulting in severe behavioural disorders, such as phobias, has a deleterious effect on self-control. Mele, however, argues that there are steps we can take to mitigate the potentially deleterious effects of visceral factors with even extreme levels of intensity, at least some of the time. Furthermore, Mele argues that the attempts we make to mitigate the potentially deleterious effects of visceral factors will themselves require that we exercise self-control. One of the conclusions we can draw from Mele is that we are able to exercise our agency with greater effect and frequency if, first, we are actively involved in the process of monitoring and evaluating our internal mental lives, and second, if we are aware of the ways in which visceral factors impact our internal states. By using the case of Wilma, Mele argues the following:

²²⁰ *Ibid.*

²²¹ *Ibid.*

Wilma, who suffers from agoraphobia, has been invited to her son's wedding in a church several weeks hence. Her long-standing fear of leaving her home is so strong that were the wedding to be held today, she would remain indoors and forgo attending. Wilma is rightly convinced that unless she attenuates her fear, she will not attend the wedding. And there is a clear sense in which she is now more strongly motivated to remain at home on the wedding day and miss the wedding than to attend it: her current motivational condition is such that, unless it changes in a certain direction, she will [stay home] [...]. Further, Wilma believes that, owing to her motivation to remain in her house indefinitely, she probably will miss the wedding.²²²

Mele argues that in the case of Wilma, there is no reason to think that she will in fact miss the wedding. Mele claims that if Wilma's desire to attend the wedding is something she values enough, it may give her the motivational strength to take active steps to actualise it, as well as to overcome the visceral effect of her phobia on her behaviour in the future. Mele says:

it certainly is conceivable that, under the conditions described, she does not have an intention to miss the wedding and, indeed, intends to do her best to reduce her fear so that she will be in a position to attend. Furthermore, there is no good reason to hold that because Wilma is now more strongly motivated (in the sense identified) to stay home on the wedding day than to attend the wedding, she also is more strongly motivated to refrain from trying to attenuate her fear than to try to attenuate it. She sees her fear as an obstacle to something she values doing, and the strength of any desire she may have to refrain from trying to attenuate her fear may be far exceeded by the motivational strength of her fear itself and exceeded, as well, by the strength of her desire to try to bring it about that she is in a position to leave her house on the wedding day and attend the wedding.²²³

Similar to the case of Wilma the phobic, Mele comments on the case of the addict, who can also take active steps to realise that in the future he will have the motivational strength to overcome the force of the visceral effect from his addiction.²²⁴ Mele's argument seems intuitively convincing here. Wilma could, for example, make appointments with a psychiatrist who specialises in helping people overcome their phobias. By doing so, she would commit herself to a course of action which would ensure that in the future she will be able to overcome the influence of these visceral factors, which currently prevent her

²²² *Ibid*, p. 109.

²²³ *Ibid*, p. 110.

²²⁴ *Ibid*, p. 99.

from being able to perform actions that either promote her values, or that track her intentions, such as going to her son's wedding. Similarly, the addict might book himself into a rehabilitation centre or join a rehabilitation program.

According to Mele, in cases of less extreme (or moderate) influence from visceral factors, their unwanted influence can be counteracted with the following divided strategies: internal and external. Mele says:

We may stem an embarrassing flow of sympathy for a character in a film by reminding ourselves that he is *only* a character. The mother who regards her anger at her child as destructive may dissolve or attenuate it by vividly imagining a cherished moment with the child. The timid employee who believes that he can muster the courage to demand a much-deserved raise only if he becomes angry at his boss may deliberately make himself angry by vividly representing injustices he has suffered at the office. These are instances of *internal* control. Many emotions are subject to *external* control as well—control through one's overt behavior. Ann defeats moderate sadness by calling her brother. Bob overcomes modest fears by visiting his coach for an inspirational talk.²²⁵

Importantly, any of the strategies suggested here, once employed, involves an exercise of self-control, and are thus useful tools for extending and improving our agency, at least some of the time.

2.7 Contingently Necessary Requirements and Constitutive Claims

So far in this chapter I have shown, through the insights from Nietzsche and Beauvoir, as well as the evidence from experimental psychology and behavioural economics, that vulnerability and difficulty are a crucial part of the story we can tell about human agency. In the previous chapter, I argued that being a human agent does not require that we exercise self-control at all times, or in all aspects of our lives. Having argued that agency is vulnerable in the face of various constraints lends further support for this claim. A drop in blood glucose levels, or the influence of external forces such as visceral factors, makes our agency vulnerable in certain circumstances, but this does not at face value give us reason to think that vulnerability and difficulty are anything more than contingently

²²⁵ *Ibid*, p. 93.

necessary requirements for the achievement of human agency. In this section, then, I argue that difficulty and vulnerability are not only contingently necessary requirements for agency but are constitutive of human agency. Following Nussbaum's lead, I argue that it is precisely because vulnerability is constitutive of human agency, that our agency can be understood as distinct from that of other sentient beings. I argue here that we must supplement the self-control view of human agency as it was described in the first chapter, since both self-control and vulnerability are constitutive of human agency.

Recall the picture sketched by Nussbaum in the epigraph to this chapter in which human beings are to be seen as: "the only beings both vulnerable and reflective, who go through life in the grip of a fear of the natural condition of their own existence, straining to understand and also to improve their condition through the reflective capacity that is also the source of much of their agony."²²⁶ On this picture, human agency (our reflective capacity) is described by Nussbaum as what distinguished us from the 'first men' and also the gods. She argues that "for many, if not for all, of the elements of human life that we consider most valuable, the value they have cannot be fully explained without mentioning the circumstances of finite and mortal existence."²²⁷ For Nussbaum, fragility and vulnerability is necessary for understanding what constitutes value, at least from our point of view. Nussbaum argues that it is only our finite and vulnerable position from which we have to enquire after the nature of our own values, and that this is the only position from which we could come to an understanding of values that would be of use or interest to us. What constitutes value for us is distinct from what constitutes value from an immortal or invulnerable position, such as would be the position of the gods. She argues that:

We are not attempting to show that an immortal existence could not have value, beauty, and meaning internal to itself [...] What we are attempting to show is the extent to which *our* values would be absent in that life [...] This is] the only perspective on value from which we can coherently proceed, in asking a question for *ourselves*: for in asking about ourselves there is not much point in asking whether a certain life seems good from the point of view of creatures that we have no chance of ever being, or

²²⁶ M. Nussbaum, 'Mortal Immortals', pp. 304-305, my emphasis.

²²⁷ *Ibid.*, p. 337.

rather creatures becoming identical to which we would no longer be ourselves.²²⁸

Nussbaum convincingly argues that what constitutes value from our point of view, and what constitutes value from the god's point of view are distinct from one another because one requires and the other lacks vulnerability. For Nussbaum this distinction is important because, she thinks, what we are interested in when asking questions about the nature of value is what constitutes value for us – that is, from our own experience and from our understanding our situation in the world as limited, finite and mortal beings. The values we would arrive at from adopting the perspective of the gods, and denying the role of vulnerability which is so central to our lives, would according to Nussbaum “be so entirely different from ours that we cannot really imagine what they would be. Nor, if we could, would they be of any immediate importance for us.”²²⁹ In support of this claim, Nussbaum argues that fragility and vulnerability are necessary for understanding what the virtues of human life are. Further, she argues that it is precisely because fragility and vulnerability are constitutive of human virtuous activity, that the virtuous activity of the gods can be distinguished from that of human agents. Nussbaum explains as follows:

Beyond this, we begin to discover that many of the virtues we prize require an awareness of the limits and needs of the human body that will be absent, as such, from a being who can never die. Moderation, as we know it, is a management of appetite in a being for whom excesses of certain sorts can bring illness and eventually death; who needs to deal with other beings similarly constituted, for whom the stakes are similarly high. Political justice and private generosity are concerned with the allocation of resources like food, seen as necessary for life itself, and not simply for play or amusement. The profound seriousness and urgency of human thought about justice arises from the awareness that we all really *need* the things that justice distributes, and need them for life itself. If that need were removed, or made non-absolute, distribution would not matter in the same way and to the same extent; and the virtue of justice would become optional or pointless accordingly [...] The closer we come to reimporting mortality – for example, by importing the possibility of permanent unbearable pain, or crippling handicaps – the closer we come to a human sense of the virtues and their importance. But that is the point: the further mortality is removed, the further *they* are.²³⁰

²²⁸ *Ibid.*, p. 340.

²²⁹ *Ibid.*, p. 340.

²³⁰ *Ibid.*, pp. 338-339.

Following the spirit of Nussbaum's argument I believe that human agency, as it is understood on the self-control view, requires difficulty and vulnerability because this is what allows us to distinguish human agency from the kind of agency which might be exhibited by supernatural, immortal and invulnerable beings such as the gods. Such beings would perhaps be capable of performing actions in the world, but the character and quality of such agency would be so different and foreign to our own that it must be considered as distinctive.

Further, according to Nussbaum's sketch, vulnerability is both the source of our reflective capacity, and a constraint on it. In section 2.5 above I argued that our agency must be seen as vulnerable in relation to internal constraints that make the kind of self-control associated with agency difficult to achieve. This difficulty and vulnerability is the direct result of the complexity of our mental lives, which provides the very conditions which make self-control possible in the first place. I think that one of the broad conclusions we can draw from this is that the constraints to our agency, as discussed in this chapter, rather than being seen as barnacles on the surface of agency, are themselves a necessary part of what it means to exercise human agency at all. Being able to exercise human agency positively requires these constraints because our agency is something that can only be achieved by working at it. Indeed, human agency is an achievement which can only be had if it can also be lost.

On the one hand, we require the presence of real and significant obstacles that threaten our agency in order to exercise our agency in the first place. On the other hand, we must also recognise the difficulty faced by any human agent in attempting to overcome those very obstacles. In recognising this, it becomes clear not only that it is difficult for everyone to be practically capable of always maintaining their agency, but also, importantly, that we understand the ease with which we so often fail in our own attempts to do so. This realisation should, I think, be profoundly disturbing for any reader who fully grasps it, for being awakened to the possibility that you yourself might not exercise your agency (at least some of the time) is bound to be unsettling. Though it may serve, at first, to unsettle us, I argue that we ought to supplement the conception of agency found in analytic accounts accordingly; our agency, I suggest, must necessarily be seen

as a struggle between control and dispersion, making us permanently vulnerable to dissolution.

Similarly, Nietzsche saw the struggle for self-control as incessant and relentless. For him, we face this struggle on an on-going and global basis. Nietzsche claims that “there results no settled state” about the psychological “resistance” that we face. For him, “the resistance must be constantly overcome.”²³¹ Here, the difficulty lies in keeping up our momentum under the strain of relentless opposition from conflicting desires and motivations, without ever arriving at a state in which we have fully mastered it all, or one in which we could rest and be absolved from the internal struggle. In this sense, Nietzsche points out that to be a human agent, who is self-controlled, is not something that we can ever take for granted about ourselves. Indeed, we must consistently fight for, and struggle with both self-control and vulnerability. In the *Twilight of the Idols*, Nietzsche claims to “understand freedom as the Romans and Venetians did, as ‘something one has and does not have, something one wants, something one conquers.’”²³² If Nietzsche is right, and I think his claims hold sway here, the very nature of our agency is necessarily vulnerable in the face of such relentless internal opposition and struggle.

2.8. Concluding Remarks

In this first part of my thesis I have advanced a view of human agency in which both self-control and vulnerability are constitutive of human agency. Human agency, I have argued, is distinct, on such a picture from the kind of agency we would attribute to non-human animals and supernatural, disembodied or invulnerable being such as gods.

In the first chapter I argued for and defended a version of the self-control view of human agency. On such a view, I argued, an agent is to be understood as someone who can muster the motivational strength to perform actions which are, by and large, informed by what might be called her ‘better judgment’. I

²³¹ R. Pippin, ‘How to Overcome Oneself’, p. 76.

²³² TI 93, in R. Pippin, ‘How to Overcome Oneself’, p. 76.

argued further that someone can only be considered to exercise their agency if they are suitable free from constraints which affect either their ability to act or their ability to exercise critical reflection, that is, to form their better judgments or plan their intentions. People whose better judgement or intentional actions are determined solely by third parties are to be seen, I argued, as lacking the kind of critical or evaluative freedom necessary to exercise agency. Finally, I have argued in the first chapter that a human agent is someone who experiences internal conflict between motivations and intentions. As such, I have argued that self-control is to be understood as the ability to muster the motivational strength to act at time t in the face of competing motivations to perform other possible actions at time t .

In the second chapter my attention turned to some of the more interesting implications of the intersection between the following: the self-control view of human agency, as described in the first chapter, Friedrich Nietzsche's somewhat controversial claims about human agency, and results of experiments in psychology and behavioural economics. While the structural and theoretical frameworks provided by Frankfurt and Watson are useful tools in explaining and understanding the mechanisms involved in the exercise of human agency, I think it is important to stress that this kind of behavioural or self-control theory is far from perfect in practice. Thus, I have argue that a more robust understanding of the nature of human agency must take into account the practical fact that not all of our attempts to modify our motivations and behaviours will be successful, and that many of our attempts will require a great amount of effort and difficulty in their achievement. Furthermore, while, as Murphy and Brown suggest, we may well be capable of restraining a cough until the end of a performance, it is far from clear that we will always be able to do so. Addictions, phobias, compulsions, reflex reactions and conditioned responses motivate people to behave in ways in which it is unclear whether there could be any way that they could really avoid doing those things. Finally, taking my lead from Freud and Nietzsche, I argued that the inner life of the human agent is typically more properly characterised as divided, and, importantly, psychological unity and stability are definitive of our agency only insofar as it is the outcome of our own efforts to achieve and preserve it, and these efforts are never complete.

Part Two: Human Agency, Personhood and Normativity

In the opening of ‘Conditions of personhood’,²³³ Daniel Dennett claims that what the necessary and sufficient conditions for personhood has not yet clearly been established in philosophical literature, saying:

One might hope that such an important concept, applied and denied so confidently, would have clearly necessary and sufficient conditions for ascription, but if it does, we have not yet discovered them. In the end there may be none to discover. In the end we may come to realize that the concept of a person is incoherent and obsolete.²³⁴

Despite this sceptical opening, Dennett goes on to outline six conditions which he claims are necessary for personhood. The necessary conditions for personhood which Dennett provides can be summarised as follows:

- (1) Rationality – a person is a rational being,²³⁵
- (2) Consciousness – a person is a conscious being,²³⁶
- (3) Recognition by the broader community of persons – “whether something counts as a person depends in some way on an *attitude taken* toward it, a *stance adopted* with respect to it”,²³⁷
- (4) Reciprocation of the personal stance – a person must be capable of adopting a personal stance towards other persons,²³⁸
- (5) Verbal communication – a person must be capable of complex verbal communication,²³⁹
- (6) Self-consciousness – a person is “conscious in some special way: there is a way in which we are conscious in which no other species is conscious”.²⁴⁰

Interestingly on Dennett’s account conditions (2) and (6) map on quite clearly to the self-control view of human agency defended in the first part of this thesis.

²³³ D. Dennett, ‘Conditions of personhood’, in Rorty, R. (ed.), *The Identities of Persons*, Berkely, University of California Press, 1976, pp. 175-196.

²³⁴ *Ibid*, p. 175.

²³⁵ *Ibid*, p. 177.

²³⁶ *Ibid*, p. 177.

²³⁷ *Ibid*, p. 177.

²³⁸ *Ibid*, p. 178.

²³⁹ *Ibid*, p. 178.

²⁴⁰ Dennett refers here explicitly to Frankfurt’s account of personhood in ‘Freedom of the Will and the Concept of a Person’. *Ibid*, p.178.

Condition (5), while contentious, could be substituted by Murphy and Brown's reflective adaptability, which as I described in the first chapter refers to "behaviours that are modified off-line by reflection using images or symbols,"²⁴¹ and thus be seen to accord with the self-control view of agency. What is most interesting, and plausible, about Dennett's suggestion is that human agency, understood in terms of the self-control view outlined in part one above, is represented as necessary but not sufficient for personhood.

Similarly, in 'Freedom of the Will and the Concept of a Person,' Harry Frankfurt not only outlines what he takes to be most distinctive about our agency, but also what he takes to be most distinctive of our personhood. Personhood, for Frankfurt, is a term which was misappropriated by P.F. Strawson, because on the Strawsonian view a person is defined merely as something which has both a body and a mind. Frankfurt argues that Strawson fails to discuss self-control or human agency in his analysis of personhood, and thus fails to describe personhood sufficiently. Contrary to Strawson then, Frankfurt outlines what he takes to be most distinctive about our personhood, outlining certain characteristics and abilities which he claims are "essential to persons" or "uniquely human", as primarily defined by our human agency. On Frankfurt's account human agency is taken to be what constitutes our personhood because it gives us the kind of control over ourselves that is thought to distinguish us from the rest of the animal kingdom. However, in his account of personhood Frankfurt, not unlike Dennett, seems to suggest that human agency, understood in terms of self-control, is necessary but not sufficient for personhood.

In the second part of this thesis, following Frankfurt and Freud, I will argue that personhood should be understood in terms of more than human agency, and that to equate human agency with personhood is to misunderstand the broader implications of the concept of personhood. The concept of a person, I will argue, is intricately linked to the concept of what it means to be a *subject*. Since our understanding of ourselves as persons involves the notion of subjectivity, I will argue that our discussion of personhood must include more complex psychological activities than those discussed when talking about human

²⁴¹ N. Murphy and W. Brown, *Did My Neurons Make Me Do It?*, p. 110.

agency. While human agency is required for personhood, persons have deeper and more psychologically complex internal conflicts to manage – and it matters, normatively speaking, how they are managed. Although self-control is central to our understanding of human agency, it is not sufficient for an understanding of what our personhood amounts to. To be a person in the fullest sense, then, I suggest is only partially constituted by self-control, and by extension, human agency.

Importantly, Frankfurt suggests in his work that while human agency, at least according to the self-control view defended in part one above, is a formal rather than a substantive notion, personhood should be understood in substantive rather than formal terms. Following this, Frankfurt proposes a normative prescription for personhood in the form of ‘wholeheartedness’ and its corollary of ‘externalisation and dissociation’. For Frankfurt through making a decisive commitment to a second order volition we become wholehearted, and in so doing we exercise our agency well.

To some extent following Frankfurt I argue that personhood is to be understood in substantive terms, and that we can thus offer normative prescriptions for personhood. However, I argue against Frankfurt’s normative prescription for personhood in the form of wholeheartedness. In the third chapter, drawing on insights from psychological horrors prevalent in the literary Gothic tradition, as well as the philosophical writings of Friedrich Nietzsche, Simone de Beauvoir and Mary Midgley, I argue that personhood normatively requires integration within a volitional complex, rather than wholeheartedness.

Following this, I argue, in the fourth chapter, for three further normative prescriptions for personhood.

First, I argue *contra* Frankfurt, that exercising our agency well lies not in what we might call wholehearted wholeheartedness, but in its opposite, ambivalence. Following Nussbaum’s discussion of value pluralism in *The Fragility of Goodness*, I argue against Frankfurt that a commitment to value pluralism entails a normative prescription for personhood in the form of ambivalence.

Second, following Nussbaum, I argue that it is an epistemic normative requirement for personhood that we recognise our own human fragility and our

vulnerability to things over which we have no control. While it seems clear that agency is necessary to our understanding of what it means to be a person, to be a person also normatively requires that a certain epistemic consideration is fulfilled: to exercise our agency well requires that a person recognise their own limits and vulnerabilities. According to Nussbaum, the ability to recognise the extent to which circumstances, events and other persons play a role “in the planning and conduct of our lives,”²⁴² contributes to our ability to live well, and should this be seen as a normative requirement for personhood. While our ability to govern and control our own actions, behaviour and lives as agents indeed separates us from the rest of the animal kingdom, I will argue that it is our uniquely human ability to recognise our vulnerability, which, in turn, defines for us as persons an epistemic normative requirement.

Finally, I argue for a normative requirement for personhood in the form of authenticity. While traditional accounts of personhood provide us with the conditions under which we gain control over our lives, to live well, I suggest, is to be engaged in the art of living. The authentic life is a life worth living because it is beautiful, and I argue that authenticity seen in this way provides a normative requirement for personhood.

²⁴² M. Nussbaum, ‘Pity and Mercy’, p. 156.

Chapter Three: Self-Governance, Wholeheartedness and Integrity

In the first part of this thesis, I argued that, on the self-control view of human agency, internal conflict is an essential feature of our mental lives as agents. I argued that because of necessary structural divisions in the mind, conflicts between desires, intentions and motivations are pervasive features in the psychological life of the human agent. Following this, I argued that part of our mental activity *qua* agents involves achieving a measure of control over these conflicts, which is what Mary Midgley refers to as the “organization of the inner crowd.”²⁴³ If we are to take this account seriously, however, I think there are further important questions we must ask about the nature of the self behind the self-control view of human agency. Most importantly: what, or perhaps more precisely ‘who,’ is doing the organising?

In the philosophical literature this question has in part been raised as an objection to Frankfurt’s hierarchical self-control view of human agency, and has been formulated in what is known as the “regress problem.”²⁴⁴ On hierarchical accounts of agency such as Frankfurt’s, self-control is achieved when there is an authenticating exchange between first order desires and second order volitions: through critical reflection, second order volitions authenticate first order volitions, and in so doing give us the requisite ownership over our actions which result from authenticated first order desires. What does the work on Frankfurt’s account is the idea that our initial responses to the world are controlled by our reflections on these responses, which are in turn controlled by our higher order responses. Essentially, this means that our initial responses to the world are authenticated by other higher order elements, which themselves exist within our psyche. The problem with this, it has been suggested, is that it is unclear what exactly authenticates those ‘higher’ elements within our psychological make-up which are responsible for authenticating the elements which we take to be our initial responses to the world (our first order desires). If our second order volitions are not themselves an authentic expression of who we are as particular persons, they must, in turn, be subject to a similar process of

²⁴³ M. Midgley, *The Ethical Primate*, p. 183.

²⁴⁴ This problem stems from Galen Strawson, but is also discussed by various other authors in the personal autonomy and freewill/determinism debate.

authentication. What we require, then, is a seemingly infinite chain of authenticating elements and processes, as Robert Noggle puts it:

No finite chain of authenticating elements can provide an account of how any element is made authentic, because no element can be the last member of the chain if every member must be authenticated by some other element.²⁴⁵

The regress problem, then, pertains to the possibility of ownership over our motivations, which seemingly requires recourse to continually higher levels of the self within the mind which endorse the more basic levels.

In addressing the regress problem, Frankfurt has supplemented his original account of self-control outlined in “Freedom of the Will and the Concept of a Person”. Frankfurt suggests that over and above the self-control required for human agency, there is a deeper and more complex psychological story that we can tell about our personhood. This more complex psychological story, for Frankfurt, involves the introduction of a psychological process which he calls ‘identification’. For Frankfurt, through the process of ‘identification’ with a second-order volition, a person exercises over and above self control, self-governance because she is in control of her higher order, as well as her first order desires. Through self-governance, for Frankfurt, we not only gain control over our mental lives, but we claim ownership of it through identification. With the introduction of this process, Frankfurt claims that his account of personhood, seen as a supplement to his account of human agency in terms of self-control, explains how we can “without arbitrariness terminate a potentially endless sequence of evaluations,”²⁴⁶ and in so doing avoid the regress problem.

In this chapter, following Frankfurt, Midgley and Freud, I argue that to be a person in the fullest sense is only partially constituted by self-control, and by extension, human agency. Personhood involves, over and above self-control, self-governance, which is reflective of a deeper kind of self-mastery.

²⁴⁵ R. Noggle, ‘Autonomy and the Paradox of Self-Creation: Infinite Regresses, Finite Selves, and the Limits of Authenticity’, in J. Taylor, (ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, Cambridge, Cambridge University Press, 2005, p. 93.

²⁴⁶ H. Frankfurt, ‘Identification and Wholeheartedness’, p. 169.

Importantly, Frankfurt suggests in his work that while human agency, at least according to the self-control view defended in part one above, is a formal rather than a substantive notion, personhood should be understood in substantive rather than purely formal terms. Following Frankfurt, in this chapter I argue that not only are persons responsible for the management of what Midgley calls the inner crowd, but it matters greatly to persons *how* internal conflicts are managed. That is, there are ways of managing the internal crowd which can be thought of as better than others, and as persons, we are thus subject to normative prescriptions for exercising our agency well.

However, in this chapter I challenge the normative prescription for personhood suggested by Frankfurt in the form of what he calls 'wholeheartedness'. In my argument against Frankfurt, I take issue with the role he sees the process of 'externalisation' or 'dissociation' playing in our becoming 'wholehearted'. For Frankfurt, in the process of achieving self-governance, while some desires are endorsed and some volitions identified with, others must be externalised, and we must dissociate ourselves from them. Drawing on insights from psychological horrors prevalent in the literary Gothic tradition, and the philosophical writings of Nietzsche and Midgley, I argue against Frankfurt's normative prescription in favour of a normative requirement for personhood in the form of integrity. Integrity, on my account, amounts to a kind of wholeness of self which Midgley describes elegantly as follows:

human freedom centres on being a creature able, in some degree, to act as a whole in dealing with its conflicting desires[...] and the conflicting desires themselves are of course not the whole story. They must belong to a being which in some way owns both of them, is aware of both, and can therefore make some attempt to reconcile them.

The more clearly the being is aware of the clash, and the more it can on occasion, distance itself from any of its impulses, feeling itself to be a whole that contains them all, the freer it becomes. This distancing does not mean taking flight to an entity immune from the conflict. Only misguided attempts at self-control are made in that way. The endeavour must be to act as a whole, rather than as a peculiar, isolated component coming into control the rest of the person. Though it is only an endeavour – though the wholeness is certainly not given ready made and can never be fully achieved, yet the integrative struggle to heal conflicts and to reach towards this wholeness is surely the core of what we mean by human freedom.²⁴⁷

²⁴⁷ M. Midgley, *The Ethical Primate*, p. 168

Contra Frankfurt, I argue that personhood requires integration of competing motivations and desires throughout one's life rather than wholeheartedness.

3.1 Self-Governance, Identification, and Wholeheartedness

Frankfurt's account of personhood is in large part formulated in response to what I discussed above in terms of "the regress problem." In attempting to answer the regress problem, Frankfurt supplemented his original account of self-control with an account of self-governance, for which the process of what he calls 'identification' is required. Recall that, on Frankfurt's account of human agency, a second order volition is formed in the volitional complex when we want a set of corresponding first and second order desires to effectively determine our will – or what we actually do in the world. When we act on a second order volition, for Frankfurt, we exercise human agency because our will has been endorsed by ourselves, and not merely imposed on us by external forces or commands. Frankfurt's process of identification supplements this account, answering the question at the heart of the regress problem, 'how are we to deal with conflicts which arise between our second order volitions themselves?' Frankfurt suggests in his account of personhood that we should resolve such conflicts by making what he calls a decisive commitment to one of our competing second order volitions. In making a decisive commitment we become what Frankfurt calls 'wholehearted'. Further, for Frankfurt, if we are unable to make a decisive commitment, we exhibit ambivalence. For Frankfurt, the ambivalent person vacillates between volitions, and this manifests in their lives as inconsistency in action as well as the inability to form a coherent picture of themselves as subjects.

For Frankfurt, the ambivalent person experiences a tension within his own volitional complex. That is, the ambivalent person can be seen to have second order volitions which contradict each other and aim to promote mutually exclusive actions on the part of that person. When a person is ambivalent, in the Frankfurtian sense, they are not able to exercise their agency well because there

is “no unequivocal answer to the question of what the person really wants.”²⁴⁸ The ambivalent person is volitionally conflicted, and thus, despite having engaged in the process of critical reflection on her first order desires, she has no one particular point of view from which to act, or which could even be seen as her point of view. In order for a person to know what *she* really wants, and thus to act in a manner reflecting *her* primary wants, she must, according to Frankfurt, “selectively identify” with one of her second order volitions.²⁴⁹ By identifying with a second order volition, the person makes that volitional desire a constitutive part of herself: she fully internalises that volition in just such a way as to make it relevantly her own.

Frankfurt claims that, by identifying with a second order volition, we need not appeal to a further higher volitional order; a person who identifies with a second order volition would find that no “further accurate inquiry”²⁵⁰ on the matter would result in her arriving at a conflicted sentiment with regard to that volition. That is to say, a person who selectively identifies with a second order volition need not make an appeal to a possible higher order volition, because, if she did, she would find the same consensus all the way up. When we have truly identified with a second order volition, we have no “endogenous desire to be volitionally different,”²⁵¹ and thus it can be said that we are exercising self-governance, that is, a deeper kind of control than the self-control required for human agency. What is distinctive about us as persons, then, is that we play an active role in the resolution of internal conflicts, which are necessary for human agency on the self-control view. While human agency requires internal conflict and constraint, a person is also someone who actively engages with and manages that internal conflict. Frankfurt claims:

If there is an unresolved conflict among someone’s second-order desires, then he is in danger of having no second order volition; for unless the conflict is resolved, he has no preference concerning which of his first-order desires is to be his will. This condition, if it is so severe that it prevents him from identifying himself in a sufficiently

²⁴⁸ H. Frankfurt, ‘Identification and Wholeheartedness’, p. 165.

²⁴⁹ *Ibid.*

²⁵⁰ *Ibid.*, p. 169.

²⁵¹ H. Frankfurt, ‘The Faintest Passion’, p. 101.

decisive way with any of his conflicting first-order desires, destroys him as a person.²⁵²

On Frankfurt's account then, a person must actively engage with the internal conflict central to exercising self-control. Moreover, if we do not engage with conflict, we become what he calls a "wanton." Frankfurt describes the "wanton" in the following terms:

The Wanton addict cannot or does not care which of his conflicting first-order desires wins out. His lack of concern is not due to his inability to find a convincing basis for preference. It is due either to his lack of the capacity for reflection or to his mindless indifference to the enterprise of evaluating his own desires and motives [...]. Since he is moved by both desires, he will not be altogether satisfied by what he does no matter which of them is effective. But it makes no difference to him whether his craving or aversion gets the upper hand. He has no stake in the conflict between them and so, unlike the unwilling addict, he can neither win nor lose the struggle in which he is engaged.²⁵³

For Frankfurt, I think correctly, a wanton is not a person precisely because he fails to play an active role in organising the inner crowd. Frankfurt puts it succinctly as follows: "When a person acts, the desire by which he is moved is either the will he wants or a will he wants to be without. When a wanton acts, it is neither."²⁵⁴ What Frankfurt is suggesting here is that, over and above controlling which of our desires move us to action, as persons, we must also *care* about which of our desires move us to action. The wanton is someone who no longer cares which direction her life takes because she does not care which of her desires she acts on. Since she no longer cares what kind of actions she takes, she no longer cares what kind of person she is, what kind of life she is leading, and what kind of impact she is having on the world. Given that she no longer cares who she is, or what she does, she is no longer any particular person at all, she is really no one.

On Frankfurt's account, then, it is only through active engagement with our internal conflicts that we can come to have the requisite ownership, required of persons, over our psychic lives. On his account, ownership of our motivations

²⁵² H. Frankfurt, 'Freedom of the Will and the Concept of a Person', pp. 15-16.

²⁵³ *Ibid.*

²⁵⁴ *Ibid.*, pp. 13-14.

is not simply given, but must be achieved. The process of achievement itself is the primary activity in which those who we consider to be persons are engaged. Through the process of critical reflection, in which our motivations are evaluated and assessed (the process required for agency on the self-control view), our motivations can become our own in a very deep sense—as constituent parts of ourselves. If who we are is constituted by the motivations which we have reflectively assessed, then, according to Frankfurt, who we are is constituted by ourselves. By being the kinds of creatures who actively engage with our internal conflict we express ourselves as particularly human subjects. Similarly, in *The Ethical Primate* Midgley's argues that persons are able to step back from the internal conflicts required for human agency and to manage and organise them. Recall that Midgley says:

human freedom centres on being a creature able, in some degree, to act as a whole in dealing with its conflicting desires [...] and the conflicting desires themselves are of course not the whole story. They must belong to a being which in some way owns both of them, is aware of both, and can therefore make some attempt to reconcile them.²⁵⁵

Further, according to Frankfurt, when we have selectively identified with one of our volitions, we not only exercise self-governance, but we can be said to be wholehearted with respect to that volition. What it means to be wholehearted, then, is that at the volitional level, a person has to be “resolutely on the side of one of the forces struggling within him [or her] and not on the side of any other.”²⁵⁶ For Frankfurt, the process of identification is important because the volitions about which we are wholehearted are sanctioned and adopted by *ourselves*, which makes them “intentional and legitimate.”²⁵⁷ According to Frankfurt:

Their force is now our force. When they move us we are therefore not passive. We are active, because we are being moved *just by ourselves*.²⁵⁸

²⁵⁵ M. Midgley, *The Ethical Primate*, p. 168.

²⁵⁶ *Ibid.*, p. 100.

²⁵⁷ *Ibid.*

²⁵⁸ H. Frankfurt, ‘Taking Ourselves Seriously’, p. 8, my emphasis.

Similarly, Frankfurt says:

Suppose we are thoroughly wholehearted both in what we are doing and in what we want. There is no respect in which we are being violated or defeated or coerced. Neither our desires nor the conduct to which they lead are imposed upon us without our consent or against our will.²⁵⁹

So, for Frankfurt, personhood requires, over and above the self-control necessary for agency, the kind of self-governance in which we constitute ourselves as particular individuals. When we are wholehearted with respect to our second order volitions we are fully directed and governed by our true or real selves and we have constituted our self through identification. By including the notions of identification and wholeheartedness, then, Frankfurt claims that he is able to avoid the regress problem in his account of personhood.

Frankfurt's suggestion that it is the person themselves who can be said to manage internal conflicts, and that it is through self-governance that the person is responsible for the formation of her 'self' as an agent, seems initially plausible because it taps into our intuitions about our own subjectivity. As human subjects, we experience our lives from our own particular point of view. Richard Moran calls this "the specifically first person authority from which, in normal circumstances, a person claims to speak his mind,"²⁶⁰ while Watson calls it "one's standpoint, the point of view from which one judges the world."²⁶¹ As human selves, we experience our own consciousness in a personal realm – the realm of our own subjective experience. The fact that we are aware of our own subjective experiences is what we most basically mean when referring to ourselves as self-aware beings. A seemingly important aspect of what it is to be a self-aware individual is that we are never uncertain of who our conscious experiences belong to – they belong to me, they are mine. When I talk about myself, then, what I refer to is, at least in part, the collection of conscious experiences which I take as belonging to me, that is to say, the conscious experiences which are *my own*. Typically, when I refer to myself I also refer to a number of things other than simply my conscious experiences: I refer to my thoughts, actions, character

²⁵⁹ *Ibid*, p. 15.

²⁶⁰ R. Moran, *Authority and Estrangement*, p. 124.

²⁶¹ G. Watson, 'Free Agency', p. 216.

traits, beliefs, goals, and preferences. The question of who I am, of who my self is, thus refers to a complex set of aspects, all of which I take as belonging, in some sense, to me. Who I am, then, can be said to be defined in contrast to what is not me, that which is outside, or other, to me.

On Frankfurt's account there are times when elements within our own psychological make-up – forces that are at work within our own psyche – arise and motivate us to act or behave in ways that we do not take to be an authentic expression of who we are. Such behaviour is not, he thinks, our own. For Frankfurt, there are a number of elements which arise within our own psyche that we experience as being “alien” to us. When someone acts on a desire or impulse that is taken to be alien in this sense, Frankfurt thinks we would describe them, as we often do, as “not being themselves,” or as acting “in spite of themselves.” Frankfurt explains, as follows:

Sometimes we do not participate actively in what goes on in us. It takes place, somehow, but we are just bystanders to it. There are obsessional thoughts, for instance, that disturb us but that we cannot get out of our heads; there are peculiar reckless impulses that make no sense to us, and upon which we would never think of acting. There are hot surges of anarchic emotion that assault us from out of nowhere and that have no recognizable warrant from the circumstances in which they erupt.²⁶²

Certainly, phobias, compulsions, reflex reactions and conditioned responses motivate people to behave in ways in which it is unclear whether there could be any way that they could really avoid doing those things. It is even less clear whether or not we would attribute these kinds of behaviours to ourselves. That is, it is unclear whether we would think of them as an expression of that which we truly want to be doing. We refer to people who act from compulsions and phobias, as victims of mental states which they themselves have had no control over. For example, we would almost never hold someone accountable for the accidents they cause while undergoing myotactic reflexes – even if such accidents might have regrettable consequences. Intuitively, at least, there do seem to be cases in which the forces at work within our psyches arise and motivate us to act in spite of our best attempts at self-governance. From time to

²⁶² H. Frankfurt, ‘Taking Ourselves Seriously’, pp. 8-9.

time, in our experience, we simply find ourselves overwhelmed by some desire or motivation which prevents us from being able to exercise self-control.²⁶³ We have all been overwhelmed by the temptations we feel the pull of most strongly, given in to the desires which are most strongly rejected by our better judgement, and found ourselves being overcome by a seemingly alien, but irresistible force from within. Moran explains this experience, as follows:

Some desires [...] may be experienced by the person as feelings that simply overcome him. They simply happen. On some occasions their occurrence may be inexplicable to him, and their inexplicability in such cases need not diminish their force. Like an alien intruder, they must simply be responded to, even if one does not understand what they're doing there or what the sense of their demand is.²⁶⁴

Experiences such as these, in which our self-governance is seen as undermined by the forces at work in our own psyche, force us, I think, to probe deeper into the claim that, through critical reflection, we are able to evaluate and assess our motivations, and, in so doing, constitute our self in its entirety. I will pick up on this concern in the following section in more detail.

3.2 Externalisation and Dissociation

What is most plausible about Frankfurt's account of personhood is that personhood is to be understood in substantive terms. On Frankfurt's account, personhood can be thought of as exercising our agency well through self-governance, and this suggestion is particularly plausible if we revise Frankfurt's

²⁶³ It is important to note that such cases are importantly distinct from cases in which an agent can be seen to exercise unorthodox self-control as discussed in chapter 1. In cases in which an agent exercises unorthodox self-control he or she exerts motivational strength in order to act (albeit an action taken against my all things considered better judgment about what I ought to do), while in the cases described here a person simply takes no action at all, he or she is simply moved by forces beyond their control (albeit forces which in a literal sense might belong to them – their neurosis or their 'tic') but, importantly, they have not exercised any motivational strength to perform an action and so have not exercised self control (not even of an unorthodox kind). These cases are distinct from one another precisely because cases of unorthodox self-control requires the exercise of motivational strength on the part of the agent, while the cases described here require nothing on the part of the agent – that is to say no exercise of motivational strength and consequently no self-control at all - the agent is, as it were, simply overcome without having done anything (they have essentially played no part in what has happened to them).

²⁶⁴ R. Moran, *Authority and Estrangement*, p. 114. See also J. Gert, *Brute Rationality: Normativity and Human Action*, Cambridge, Cambridge University Press, 2004.

views on what self-governance amounts to. Importantly, on Frankfurt's account, since personhood is to be understood in substantive terms, he argues that we can offer a normative prescription for personhood in the form of 'wholeheartedness'. In this section I take issue with Frankfurt's normative prescription for wholeheartedness which he argues requires 'externalisation and dissociation'. For Frankfurt, in the process of achieving wholeheartedness, while some desires are endorsed and some volitions identified with, others must be externalised, and he suggests that we ought to dissociate ourselves from them. In this section I problematise Frankfurt's suggestion that we ought to practice dissociation and externalisation, and ultimately, as I argue in what remains of this chapter, reject Frankfurtian wholeheartedness in favour of a normative prescription for personhood of integrity.

Above I discussed Frankfurt's idea that there are times when elements within our own psychological make-up-- forces at work within our own psyche - arise and motivate us to act or behave in ways that we do not understand as authentic expressions of who we really are. That is, Frankfurt suggests, we think that such behaviour is not our own. Some of our desires and volitions, as Frankfurt suggests, are only to be considered as ours in the same strictly literal way that a nervous tic is considered ours; they occur within our bodies but we do not take them to be an expression of what we, ourselves, actually desire or volitionally will. Recall that Frankfurt says: "Sometimes we do not participate actively in what goes on in us. It takes place, somehow, but we are just bystanders to it."²⁶⁵ In a similar vein, Watson explains:

One measure of the strength of a desire is their capacity to claim one's consciousness, direct one's fantasies, break one's concentration on other things. One finds it difficult to keep one's mind on one's work because one keeps thinking of one's lover, or of the chocolate cake in the pantry, or of the cigarettes at the market. The objects of these desires tend to demand or dominate one's attention, *despite oneself*.²⁶⁶

And Watson says again:

²⁶⁵ H. Frankfurt, 'Taking Ourselves Seriously', pp. 8-9.

²⁶⁶ G. Watson, *Agency and Answerability*, pp. 71-72, my emphasis.

Just as the bouncer can force you out of the room contrary to your will, so your appetites and impulses might lead you where you do not 'really' want to be.²⁶⁷

On such a picture, these alien or rogue desires and volitions are therefore external to us *because* we do not accept them as legitimate reasons for performing acts that we would otherwise consider true expressions of who we actually are. For this reason, Frankfurt claims that "even if an externalized desire turns out to be irresistible, its dominion is merely that of a tyrant. It has for us no legitimate authority."²⁶⁸ He says, in a passage from "Taking Ourselves Seriously," the following:

Let us suppose that a certain motive has been rejected as unacceptable. Our attempt to immunize ourselves against it may not work. The resistance we mobilize may be insufficient. The externalized impulse or desire may succeed, by its sheer power, in defeating us and forcing its way. In that case, *the outlaw imposes itself upon us without authority, and against our will.*²⁶⁹

On Frankfurt's view, we react with hostility to encounters with these alien forces within us. We want to disown and reject them because we do not think that they belong to us, and we wish to regain a measure of self-governance by ridding ourselves of them. Similarly, Marilyn Friedman explains her experience of this phenomenon, as follows:

There are times for example when I feel wholly consonant with my motivations, fully satisfied with the choices to which they may move me; at other times, my motivations feel alien and, although they do not cease to move me, nevertheless, I want to be free of them as I would want to be free of a fever, an ache, a disturbing or painful condition that causes me grief and is not, in any way, my "self."²⁷⁰

On this view, it is our reaction to this experience, the hostility and sincerity with which we disavow the alien forces erupting within us, which highlights one of

²⁶⁷ *Ibid.*, p. 62.

²⁶⁸ H. Frankfurt, 'Taking Ourselves Seriously', p. 10.

²⁶⁹ *Ibid.*, p. 14, my emphasis.

²⁷⁰ M. Friedman, 'Autonomy and the Split-Level Self', *The Southern Journal of Philosophy*, vol. XXIV, no. 1., 1986, pp. 19-35, p. 27.

the ways in which we come to experience the self as other. J. David Velleman describes this reaction as illuminating a kind of self-division in which the fundamental divide between the self and the other is internal to us. That is, it occurs within our own psyche. He says:

Among the goings on in a person's body, some but not others are due to the person in the sense that they are his doing. When he distinguishes between those which are his doing and those which aren't, he appears to do so in terms of their causes, by regarding the former but not the latter as caused by himself. Yet even when he disowns them, he ends up disowning parts of his own body and mind, as if the boundary between self and other lay somewhere inside the skin.²⁷¹

We might find ourselves asking the following question: “[c]ould that really have been me who thought or felt or did those things?” Our reply is only that it “could not really have been me,” because it is not, “in any way, *my* self.”²⁷²

The natural question arising here is what are we to do with these rogue elements? Or perhaps, more properly, how are we to manage this internal conflict? In answer to these questions, Frankfurt provides the following normative prescription for personhood: we must dissociate ourselves from certain unwanted or unacceptable desires and volitions in order to be wholehearted and thus exercise self-governance. According to Frankfurt, when we encounter any rogue elements within our own psyche we must strive to “dissociate ourselves from them,” rather than attempting to incorporate them into our conception of ourselves, saying:

That is, we dissociate ourselves from them and seek to prevent them from being at all effective. Instead of incorporating them, we externalize them [...]. They are outlawed and disenfranchised. We refuse to recognise them as grounds for deciding what to think and what to do [...] even if an externalized desire turns out to be irresistible, its dominion is merely that of a tyrant. It has, for us, no legitimate authority.²⁷³

²⁷¹ J. Velleman, *Self to Self*, p. 7.

²⁷² M. Friedman, ‘Autonomy and the Split-Level Self’, p. 27, my emphasis.

²⁷³ H. Frankfurt, ‘Taking Ourselves Seriously’, p. 10.

In a similar vein, Watson argues that in cases where we come up against rogue desires that undermine our self-governance, we are to resist and reject such elements as being part of our true or real self. According to him:

The possibility I have in mind is rather that what one is disposed to say or judge is temporarily affected by the presence of the desire in such a way that, both before and after the “onslaught” of the desire, one judges that the desire’s object is worth pursuing (in the circumstances) whether or not one has the desire. In this case one is likely, in a cool moment, to think it a matter for regret that one had been so influenced and to think that *one should guard against* desires that have this property.²⁷⁴

Both Frankfurt and Watson suggest that what is central to the notion of our true or real self is *who we take ourselves to be*, regardless of the presence of rogue or unwanted desires, motivations or volitions. They also agree that we must actively externalise all aspects of ourselves that are contrary to who we take ourselves to be, for such aspects are not legitimately part of our self-constitution. If we are unable to externalise these rogue motives and impulses, which do not stem from our real selves, Frankfurt and Watson argue that our self-governance is undermined. According to Frankfurt and Watson, through dissociation and externalisation, we declare that these rogue elements are external to us, and, as such, we declare that certain elements of ourselves are not constitutive of our true or real selves. For them, by rejecting and resisting the allegedly rogue elements of our mental empire, we take responsibility for determining the boundaries of our true or real selves. In short, we declare that this is precisely where the boundary between self and other lies.

In her paper ‘Autonomy and the Split-Level Self,’ Friedman explains the move made by Frankfurt and Watson when she says that “[s]plit level self theorists tell us, in varying ways, that critical reflection [rational deliberation and evaluation] is more ‘truly of the self’ than unassessed motivation.”²⁷⁵ Friedman further explains this move as follows:

It is also sometimes claimed that the higher level self which does the critical reasoning [or makes rational deliberations and evaluations], is

²⁷⁴ G. Watson, ‘Free Agency’, p. 214, my emphasis.

²⁷⁵ M. Friedman, ‘Autonomy and the Split-Level Self’, p. 23.

the “true” self, a status which is evidently supposed to be what guarantees that critical assessment makes motivation truly self selected, and hence, the basis of autonomous choice.²⁷⁶

Importantly, she goes on to ask what exactly such theorists mean when they claim that some element of the psyche is “more truly of the self” than any other element of that whole psyche. I agree with Friedman when she claims that the notion of isolating certain elements in our own psychological make-up, and referring to these as the only relevant factors when describing our identity, seems exceedingly arbitrary:

it is unclear how one part of the self could be more truly that self than any other part. And, in sheer ontological terms, it is far from clear how a part which is less than the whole of something could nevertheless be more truly that (whole) thing than the whole thing itself.²⁷⁷

Moreover, I agree with Friedman’s claim that the very process required for determining the elements, which form the true self, relies on an assessment which can only be made in light of an understanding that there be some other prior, whole self. She explains, as follows:

The self who recognizes which part is its true self could only be the original whole self, and this whole self continues to remain conceptually distinct from the constituent part which is regarded as its ‘true’ self. [...] [I]t is the whole self, not the ‘true’ self, which is doing the “identifying” with that constituent regarded as the ‘true’ self.²⁷⁸

I think that Friedman’s line of questioning here is worth pursuing. Further, I think it is worthwhile to question the fundamental assumption about the constitution of the true or real self which lies behind accounts of self-governance, such as those put forward by Frankfurt and Watson: we must ask how any one part of the self could be considered to be a more authentic expression of that self than the whole of that self.

Contra Frankfurt, Freud and Mary Midgley (following Jung), have “suggested how little we understand about our own motives and what deep

²⁷⁶ *Ibid*, p. 22.

²⁷⁷ *Ibid*, p. 28.

²⁷⁸ *Ibid*, pp. 28-29.

conflicts perturb even those motives of which we are more or less conscious.”²⁷⁹ This suggestion is again intuitively plausible. In our experience, who we are and who we most truly take ourselves to be, seems to not always map onto what we take our better judgement to be. Typically, we find ourselves in a number of situations in which we find it difficult to know who we most truly are. Indeed, in our reflection, there are many past situations that reveal to us precisely how misinformed we were about who we took ourselves most truly to be.

The writings of philosophers and psychologists, such as Nietzsche, Freud and Lear, show “that there are deep currents of meaning, often crosscurrents, running through the human soul which can at best be glimpsed through a glass darkly.”²⁸⁰ Following these thinkers, it seems that our ability to exercise self-governance is far more complex in psychology than is suggested in the analytic philosophical literature. Recall from my discussion of Freud’s account in the first chapter, Freud sees the human mind as divided into the structures of the ego and the id. This psychological picture served us well when discussing the mental life of a human agent, but it is not enough to give us the full psychological story of a person. Crucially, Freud saw the human psyche as divided in a further, important way. Freud also saw the human mind as also being divided between the conscious and unconscious aspects. This division in the psyche allows Freud to provide a richer psychological account of personhood in which personhood involves deeper psychological complexity than what is required on the self-control view of human agency.

That Freud saw the human mind as also being divided between the conscious and unconscious aspects, has perhaps been the most crucial insight into the human mind in the history of psychology, and forms the backbone of the psychoanalytic tradition that has developed as an influential branch of psychology over the last century. Anthony Elliott explains that, from the psychoanalytic perspective, this internal boundary reveals a hidden, unconscious self that lurks within the psyche beneath the superficial appearance of a unified, rational and conflict-free self. According to Elliott:

²⁷⁹ M. Midgley, *The Ethical Primate*, p. 21.

²⁸⁰ J. Lear, J. Lear, *Open Minded: Working Out the Logic of the Soul*, Cambridge, Massachusetts, Harvard University Press, 1998, p. 28.

Dismantling the notion of an essential unity to the self, psychoanalysis posits a split at the centre of the psyche between consciousness of self and that which is unconscious. Lurking behind all forms of self-organization – that is, our day to day fashioning of self-identity – there lies a ‘hidden self’, a dimension of subjectivity that is cut off from self-knowledge.²⁸¹

On the psychoanalytic picture, the unconscious or hidden self has a significant influence on the formation and governance of the conscious aspects of the self. Importantly, the unconscious self is often in conflict with conscious aspects of the self. As Elliott explains:

Significantly, this hidden self, however we may choose to act or express ourselves, constantly disrupts and outstrips our intentions [...].²⁸²

Drawing on Freud’s suggestion that unconscious motivations influence much of our behaviour, psychoanalytic theorists dismiss the assumption that people essentially act in “rational and transparently explicable ways,” focusing rather on understanding the fact that “people often act in bizarre ways, ways which cause pain to themselves and others, ways which puzzle even the actors themselves.”²⁸³ Through the development of the broadly psychoanalytic picture, what emerges is a far more complex picture of our psyche, which is largely characterised by higher order conflict and is far from transparent to itself.

In what follows in this chapter, Following Midgley and Freud, I challenge Frankfurt and Watson’s normative prescription for personhood in the form of externalisation and dissociation, using late gothic novels as a springboard for my discussion. I argue that, while self-governance – over and above self-control – is required for personhood, following Frankfurt’s normative prescription for personhood in the form of wholeheartedness is problematic because it leads to a psychologically imbalanced self. Personhood, I argue, requires psychological stability and unity insofar as it relates to our subjectivity, and thus, *contra* Frankfurt and Watson, I will argue in what follows in this chapter for a normative prescription for personhood in the form of integrity.

²⁸¹ A. Elliott, *Psychoanalytic Theory*, p. 10.

²⁸² A. Elliott, *Psychoanalytic Theory*, p. 10.

²⁸³ J. Lear, *Open Minded*, p. 25.

3.3 Gothic Subjectivity

In *The Fragility of Goodness*, Martha Nussbaum argues that literary works provide a commanding source from which to refine our understanding of human experiences, or of our subjectivity, which she takes to be defined, in part, by interpersonal and intrapersonal clashes. She suggests that we can refine our understanding of our subjectivity by turning to literary works because they reveal to us, in a particularly poignant way, “the vulnerability of human lives to fortune, the mutability of our circumstances and our passions, [and] the existence of conflicts among our commitments.”²⁸⁴ Fitting with her primarily Aristotelian account of personhood and subjectivity, Nussbaum’s study focuses on the literature, in particular the tragedies, of the ancient Greeks. In my discussion of the motifs and themes of gothic literature, particularly of the late nineteenth and early twentieth centuries, I will, to some extent, rely on the example set by Nussbaum in order to delve into the psychological explorations of subjectivity in this genre.

In this section, I will argue that gothic fiction provides us with insight into the psychological aspects of our subjectivity, through which we can explore the issues of self-governance and personhood. Gothic fiction in many ways gives literary form to philosophically interesting questions about the nature of our own subjectivity, namely the nature of the relationship of the self with itself. Contra Frankfurt and Watson’s normative prescription that we ought to achieve self-governance through externalisation and dissociation (outlined in the previous section), I argue in what follows in this thesis that the picture of self-governance that Gothic authors advance highlights the psychological instability which would result from following this prescription. On the gothic picture of self-governance, the self is to be understood as Midgley puts it, a whole-self, or as I argue a self with integrity. Integrity, I argue then, can be seen as a normative requirement for personhood.

²⁸⁴ M. Nussbaum, *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*, 2nd edn., Cambridge, New York, Cambridge University Press, 2001, p. 13.

3.3.1 Late Gothic Fiction and Subjectivity

In this century, academic philosophy, as much as psychology, has been reluctant to pay much attention to the shadow-side of human motivation. It has not occupied itself with the agonizing question 'can it really have been I who did that?' or with the genuine clash of reasons for answering yes or no to it. Nor has it dealt much with the still more startling division of the self into two or more selves and shadows embattled factions which marks the process of temptation.²⁸⁵

It may at first seem strange to incorporate an analysis based on Gothic fiction in my discussion of the nature of personhood and subjectivity. I think that Gothic fiction has, to a large extent, been widely misinterpreted because of its misleading associations with the horror genre. In literary criticism circles, however, Gothic fiction has recently been defended against our misgivings about its ties to the horror genre. As literary critic Gary Thompson explains:

Preternaturalism has, of course, been a source of annoyance to some critics of the Gothic; and it does, indeed, require a strong palate to accept all the bleeding portraits, animated skeletons, lycanthropes, rattling chains, and vampires that infest Gothic literature, especially the older novels. But the artistic incorporation of the preternatural should not, in itself, form a barrier to critical appreciation.²⁸⁶

Robert D. Hume has argued that when we look at Gothic fiction in particular, what we are primarily analysing are textual representations of aspects of human subjectivity. He claims that one of the major Gothic concerns "might grandiosely be called a psychological interest [...] [where] there is a considerable amount of concern for interior mental processes."²⁸⁷ Following his famous study in defence of the psychological and philosophical importance of the genre, literary criticism from the 1960s and onwards has systematically attempted to analyse Gothic fiction in terms of our subjectivity by exploring the extent to which it seeks to

²⁸⁵ M. Midgley, *Wickedness*, p. 119.

²⁸⁶ G. Thompson, *The Gothic Imagination: Essays in Dark Romanticism*, Pullman, Washington State University Press, 1974, pp. 17-18.

²⁸⁷ R. D. Hume, 'Gothic vs Romantic: A Revaluation of the Gothic Novel', *PMLA*, vol. 84, no. 2, 1969, p. 283.

unmask psychological dread and fear by venturing into “the dark night of the irrational.”²⁸⁸

Elizabeth Napier has challenged the view put forward by Hume and his followers, with particular respect to the early Gothic novels of the late eighteenth century. She claims, first, that the psychological explorations in the early Gothic are one-dimensional in their refusal to deal with morally mixed characters, since in these novels the “absolute polarization of good and evil necessitates that virtuous and erring characters finally remain permanently apart.”²⁸⁹ However, there is a marked shift in late Gothic fiction of the late nineteenth and early twentieth century, which distinguishes them from their early Gothic predecessors (that they perhaps more accurately outgrew). Napier’s first conclusion about the early Gothic is that it is a mistake to think that we can gain significant psychological insight from it because of the lack of depth in the early Gothic characters’ personalities. However, it is precisely this depth of character, or self, that the late Gothic seeks to explore. By delving into the dark and dingy corners of the labyrinth at the heart of the human psyche, late Gothic novels present us with morally mixed or ambiguous characters, who are disturbingly and ambiguously internally divided. For this reason, I see the late Gothic novel as a fitting setting in which to explore those interesting philosophical questions about the constitution of a mode of experience in which we encounter the self as other.

Late Gothic, or what Richard Davenport-Hines calls the “gothic revival,” is seen primarily as a reaction to the Enlightenment’s “emphasis on the need for rationality, order, [and] sanity [...] in which human happiness and achievement rested on the mastery of passion, and on calm, confident regulation”²⁹⁰ of potentially conflicting aspects of the self. Arguably, the early Gothic authors recognised that this assumption rested on “the opposing half-truth that humankind needs passion and fear,”²⁹¹ but failed to fully explore the implications

²⁸⁸ M. Miyoshi, *The Divided Self: A Perspective on the Literature of the Victorians*, New York, New York University Press, 1969, p. 5.

²⁸⁹ E. Napier, *The Failure of the Gothic: Problems of Disjunction in an Eighteenth Century Literary Form*, Oxford, Oxford University Press, 1987, p. 11.

²⁹⁰ R. Davenport-Hines, *Gothic: 400 Years of Excess, Horror, Evil and Ruin*, London, Fourth Estate, 1998, p. 3.

²⁹¹ *Ibid.*

of this for human psychology and subjectivity in their literary treatments of this realisation. In contrast, the literature of the late Gothic authors is concerned with irrationality, the forces of the unconscious, and the real and pressing existence of a tension between disparate and conflicting forces at work within the person themselves, which Davenport-Hines aptly describes as “the power of inward goblins to torment one’s psyche.”²⁹² Unlike their predecessors, late Gothic authors seek to explore literary characters that provide us with cases that exhibit a failure of agency or a breakdown of personhood, in which the characters’ thoughts, actions and behaviours are not a reflection of who they really think they are, or who they most truly are. In these cases, characters encounter the other within themselves. Recently, in support of this argument, Julian Wolfreys explains the following:

If there is a transition in the nature of the gothic from the end of the eighteenth century to the middle years of the nineteenth century, it is marked by an inward turn [...]. In writing of the nineteenth century which manifests a gothic turn, there is an embrace of the uncanny other within ourselves rather than a displacement or projection on to some foreign or distant other.²⁹³

Late Gothic fiction, then, marks a transition in horror stories from fearful transcendental explorations, which involve supernatural realms and otherworldly demons, to fearful internalised explorations, in which fears and anxieties are generated by and within the self. I take late Gothic literature as providing a powerful source for refining our understanding of subjectivity by “descending [...] into the darker and ‘lower’ realms of human experience, where one realises in a fairly visceral way both the precariousness of life and the existence of powers greater than oneself.”²⁹⁴ In doing this, Gothic horror intends to shock us by asking us to question our own experiences as subjects. In Gothic psychological horror, the precariousness ambiguity of life is specifically revealed to us in the deeds undertaken by the hero-turned-villain at the centre of the tale. Through his dastardly and violent deeds, we are made poignantly aware not only

²⁹² *Ibid.*

²⁹³ J. Wolfreys, ‘I Wants to Make Your Flesh Creep’, in J. Wolfreys, and R. Robbins, (eds.), *Victorian Gothic: Literary and Cultural manifestations in the Nineteenth Century*, Basingstoke, Palgrave Macmillan, 2000, p. 31.

²⁹⁴ *Ibid.*

of human physical mortality and frailty (in the torture and death of his victims), but also of the hero's psychological frailty as we see him give in to temptation, indulgence and lose the internal struggle in the face of internal conflict and division. Having identified with the hero, we are, in turn, made aware of our own frailty and the existence of psychological powers that are sometimes overwhelmingly beyond our ability to control or remove from our psychological constitution. Ultimately, the drama and horror in the Gothic narratives involve shocking us because they demand that we recognise a very difficult and painful truth about ourselves – that we are less psychologically unified than we typically take ourselves to be, and that our own attempts at psychological unity and stability could fail just as easily as the hero's. Late Gothic fiction thus provides us with stories generated by the uncanny transcriptions of a fragile human psychological condition with which we identify, from our own lived experience, but that we often struggle to fully comprehend or embrace. In the following section, I will argue that this is particularly prominent in Gothic novels which explore the motif of the double.

3.3.2 Internal Division, the Double, Whole Selves and Integrity

The image of the double – the Doppelgänger, or second self, the mirror image, the Other who is also oneself – tracks, haunts, or shadows cultural production in the nineteenth century, and on into the twentieth.²⁹⁵

In this section I will turn my attention to a literary motif which undoubtedly seems to explore the issue of self-division: the double, otherwise known as the doppelgänger, the second or shadow self, or, sometimes, the *alter-ego*. Stories that centre on the doppelgänger, in late Gothic fiction, present main characters and their doubles as psychologically divided and fractured aspects of the same self. The motif thus illustrates an understanding that the boundary between the self and the other also appears to be applicable to divisions within the psyche itself.

²⁹⁵ D. Jones, *Horror: A Thematic History in Fiction and Film*, London, Arnold, 2002, p. 103.

In its dramatisation of this view of the double self, late Gothic fiction presents us with a main character who struggles to come to terms with their own duplicity. The character ends up in a disastrous, nightmarish situation marked not by fear of the other, but by fear of having to embrace the other within. The drama in the Gothic begins to unfold when the once hidden duplicity of self is revealed: the double typically emerges by appearing in a mirror, portrait, or encounter with another person who is, in appearance, identical to the self. The Gothic drama reaches its climax as the profound power and influence of the once hidden aspects of the self on the individual are realised. The intentionally unsatisfying resolution of these stories is achieved when it is realised that the conflict within, and fragmentation of the self, can only be circumvented or controlled through the murder of the double, which ultimately results in death for both parties. Drawing on this motif in Gothic fiction, I argue, in this section, that Frankfurt's normative prescription, in the form of wholeheartedness is problematic because it relies on the process of externalisation and dissociation.

With the development of psychoanalytical literary investigations, starting with the seminal work by Otto Rank,²⁹⁶ our understanding of the motif of the double has shifted. Once interpreted as an author's depiction of their longing for a different existence, the interpretation of the double has evolved as a manifestation of the subconscious desire of an author to "lend imagery to a universal human problem – that of the relation of the self to the self."²⁹⁷ Late Gothic authors such as Oscar Wilde, Edgar Allan Poe and Robert Louis Stevenson (to name but a few), explore this "universal human problem" most explicitly through the concept of the doppelgänger – a sinister double of the individual who represents the shadow side of human motivation. In their literary explorations, the relationship between the self and itself is marked by extreme tension and radical division. As William Patrick Day explains, "the double is a central motif in the genre; [where] the individual is not one self, but two."²⁹⁸ The

²⁹⁶ O. Rank, *The Double: A Psychoanalytic Study*, Tucker, H., trans., (ed.), United States of America, The University of North Carolina Press, 1971.

²⁹⁷ H. Tucker, 'Introduction' in *Ibid*, pp. xv-xiv.

²⁹⁸ W. Day, *In the Circles of Fear and Desire: A Study of Gothic Fantasy*, Chicago and London, The University of Chicago Press, 1985, p. 6.

protagonists and their doubles are essentially characterised as psychologically distinct and fractured aspects of the same self. As Davenport-Hines argues, the late Gothic authors reject a “bourgeois” understanding of the self as “stable, abiding and continuous, requiring the assertion of one true cohesive inner self.”²⁹⁹

Using preternaturalism as a springboard, Gothic fiction provides us with literary representations of internal conflict, as well as the effect of such conflict on our subjectivity. As Peter Garrett explains, “the uncanny events and effects of the Gothic already” work to “estrangle us from the familiar.” For him, the “reflexivity” of these events and effects “is always linked with the problematic” and internal “relations of subjectivity.” That is, Garrett continues, the Gothic character’s “self-consciousness [is] always in tension with the forces of the unconscious.”³⁰⁰

Similarly drawing on Freud’s picture in which the human mind is crucially divided into conscious and unconscious aspects, the psychoanalytic perspective maintains that the internal boundaries of the self reveal a hidden, unconscious self, which lurks in the psyche, beneath the superficial appearance of an Enlightenment picture in which the self is unified, rational and free of conflict. Moreover, on this psychoanalytic picture, the unconscious or hidden self has a significant influence on the formation and governance of the conscious aspects of the self and, importantly, is often discordant with those conscious aspects.

This psychoanalytic picture of the self is clearly at play in the development of the motif of the double in late Gothic fiction. Indeed, many literary critics have drawn parallels between psychoanalysis and Gothic fiction, often providing psychoanalytic readings of Gothic novels. The fact that both psychoanalysis and the Gothic are reactions against the Enlightenment picture of the self, gives us reason to think that many fruitful comparisons are to be made between them. As Michelle Massé puts it, “[i]f the Gothic can be said to influence psychoanalysis, psychoanalysis in its turn illuminates the Gothic explicitly and

²⁹⁹ R. Davenport-Hines, *Gothic*, p. 7.

³⁰⁰ P. Garrett, *Gothic Reflections: Narrative Force in Nineteenth-Century Fiction*, London, Cornell University Press, 2003, p. 9.

implicitly.”³⁰¹ In the late Gothic, we see that the ultimate assumption of the Enlightenment – the unity of the self – is challenged through the motif of the double, in which there is a shift from a portrayal of the other, as something contrasted to some entity or person other than the self, to a portrayal in which the dialogues between the self and other turn out to be colloquies within a fractured and conflicting self.³⁰²

The motif of the double, as expressed in late Gothic fiction, particularly and effectively lends itself to these kinds of psychoanalytic interpretations of self-division. The motif is often employed in these texts through a device such as a mirror or portrait, which reflects the self as a character’s double or other. This reflection is one that is (or would be) otherwise *hidden* from their view. In *Dr Jekyll and Mr Hyde*, for example, Dr Jekyll writes the following in his “Full Statement of the Case:”

There was no mirror, at that date, in my room; that which stands beside me as I write, was brought there later on and for the very purpose of these transformations [...] [W]hen I looked upon that ugly idol [Mr Hyde] in the glass, I was conscious of no repugnance, rather a leap of welcome. This too was myself.³⁰³

Similarly, in Wilde’s tale, the painted portrait of Dorian Gray functions, perhaps even more centrally, “as an iconographical establishment of difference”³⁰⁴ in which the self is reflected back to itself as other. Like psychoanalytic theory, the Gothic picture of the self is one in which we do experience ourselves as internally divided. The suggestion of Gothic literature is that this is not experienced as an accidental feature of a person, but rather an essential feature. Moreover, in revealing to the protagonists those aspects of themselves which would otherwise be hidden from view, these Gothic tales bring the experience of confrontation between disparate aspects of the self to the forefront of their narratives.

If we, like the characters in Gothic novels, experience ourselves as internally divided, and at times experience conflict between the various aspects

³⁰¹ M. Massé, ‘Psychoanalysis and the Gothic’, in Punter, D., (ed.), *A Companion to the Gothic*, United States of America, Blackwell, 2001, p. 230.

³⁰² R. Jackson, *Fantasy: The Literature of Subversion*, London, Methuen, 1981, p. 108.

³⁰³ R. Stevenson, *The Strange Case of Dr Jekyll and Mr Hyde*, Danahay, M., (ed.), Canada, Broadview Press Ltd., 2000. p. 70.

³⁰⁴ R. Jackson, *Fantasy*, p. 45.

of our psyche, it seems only natural to ask how we ought to deal with this conflict and division. In this section, I will return to that which I mentioned earlier in this chapter: our reaction to the experience of internal conflict and division, and its effect on our self-governance. Recall the intuition that when we encounter alien forces within us, we react to them with hostility: we want to disown and reject them because we do not think that they belong to us. Following this intuition, Frankfurt claims that when we encounter any rogue or alien elements within our own psyche we must strive to “dissociate ourselves from them,” rather than attempt to incorporate them into our conception of ourselves. According to Frankfurt, then, through dissociation we declare that these rogue elements are external to us. As such, we declare that certain elements of ourselves are not constitutive of our true or real selves. By rejecting and resisting the allegedly rogue elements of our mental empire, according to Frankfurt, we take responsibility for determining the boundaries of our true or real self, and in doing so gain the kind of self-control essential to self-governance and personhood.

The question of how we should deal with this conflict and division is also central to Gothic literature: Alison Holland explains that “the most characteristic energies of the Gothic novel concern the impossibility of restoring to their original oneness characters divided from themselves.”³⁰⁵ Rather than an assumption of essential unity to the self, the Gothic double reveals – as Judith Halberstam explains – the ‘essential duplicity and potential multiplicity’ of the self.³⁰⁶ In Stevenson’s tale, Dr Jekyll attempts to rationalise his internal conflict by universalising his condition in order to minimise his culpability for leading a life of duplicity. He argues that his divided life was only a more extreme instance of an eternal human dilemma. As Dr Jekyll proclaims, “man is not truly one, but truly two [...] and I hazard a guess that man will ultimately be known for a mere polarity of multifarious, incongruous and independent denizens.”³⁰⁷ The internal conflict between the various parts of the psyche is then, as we have seen, central

³⁰⁵ A. Holland, *Excess and Transgression in Simone de Beauvoir’s Fiction: The Discourse of Madness*, Farnham, Ashgate Publishing Ltd., 2009, p. 35.

³⁰⁶ J. Halberstam, *Skin Shows: Gothic Horror and the Technology of Monsters*, Durham and London, Duke University Press, 1995.

³⁰⁷ *Ibid*, p. 76.

to the Gothic picture of fractured subjectivity. Importantly, however, this inner conflict is characterised as threatening, especially because, “[p]roverbially, an encounter with the double portends the death of the self,” and “closely allied to this are fears of the portrait, or of the mirror, as displaying a corrupt or evil image of the self.”³⁰⁸

The basic plots of various novels exploring the above-mentioned themes, such as that of *Dr Jekyll and Mr Hyde* or *Frankenstein*, have found fame in non-literary circles as children’s stories or have been used as descriptively loaded catchphrases. People often mistakenly refer to Frankenstein as the Golem who was sparked to life by lightning, and use *Dr Jekyll and Mr Hyde* as an almost Homeric epithet when referring to people with a split personality or bipolar disorder. The problem with these non-literary adoptions is that they are prone to provide misinterpretations that lack the subtle nuances that are fundamentally important to the texts themselves. This is particularly telling in the case of *Dr Jekyll and Mr Hyde*, for the novel itself is neither intended as a guide-book for diagnosing mental disorders, nor as a warning about how monstrous we may well become when we give into the temptation of our sinful natures. What is most apparent, and indeed most alarming, is that the warning from Stevenson is not that we all have the potential to become Mr Hyde when we give in to temptation. Rather, the warning is that if we, like Dr Jekyll, attempt to wholly separate ourselves from the Mr Hyde, who dwells in each of us, we are simply fracturing ourselves rather than achieving integrity and this can only be detrimental for our selves. In order to understand the ways in which this warning is relevant to the notions of self-governance and personhood and, in particular, the problems encountered in both Frankfurt’s and Watson’s accounts, we will need to examine more carefully the exact manifestation of this warning in the novel itself.

From the outset of the novel, Dr Jekyll is aware that he is a deeply divided individual. I suggest that his division, rather than occurring because of a lack of self-control, should be seen as a manifestation of, in Frankfurt’s grammar, the second order volitional level. Using Frankfurt’s terms, Dr Jekyll can be described as having two radically different and diametrically opposed higher order or

³⁰⁸ D. Jones, *Fantasy*, p. 103.

evaluative selves, both with which he seems to identify. To this effect, Dr Jekyll claims the following:

though so profound a double-dealer, I was in no sense a hypocrite, both sides of me were in dead earnest; I was no more myself when I laid aside restraint and plunged in shame, than when I laboured in the eye of day.³⁰⁹

Dr Jekyll, then, is what Frankfurt would call ambivalent. That is, Dr Jekyll experiences a volitional division that prevents him from having a unified identity, which, in turn, and according to Frankfurt, prevents us from thinking that there is “a certain truth about him at all.”³¹⁰ Stevenson, however, in my opinion, seems to question Frankfurt’s assertion when he wrote the following claim which is uttered by Dr Jekyll in the novel:

I saw that, of the two natures that contended within the field of my consciousness, even if I could rightly be said to be either, it was only because I was radically both.³¹¹

With regards to the notion of ambivalence, however, Stevenson seems to agree with Frankfurt that such a division would at least hinder the ambivalent individual’s ability to achieve certain of his goals or to follow a single overarching life plan at all. Dr Jekyll himself admits, with regards to his internal division, that “this incoherency of my life was daily growing more unwelcome.”³¹²

In an attempt to explore the complete polarity of selves, which Dr Jekyll experiences within himself, we see Dr Jekyll’s interest turn to performing experiments which aim to utterly separate the two sides of his nature. Dr Jekyll attempts to escape his dual nature, which Frankfurt would call his ambivalence, by externalising certain aspects that are associated with only one side of his nature. In order to do so, he attempts to wholly separate one half of his nature and thereby create what he considers to be an entirely separate person, namely Mr Hyde. In the manifestation of Mr Hyde as a separate individual to Jekyll,

³⁰⁹ R. Stevenson, *The Strange Case of Dr Jekyll and Mr Hyde*, p. 76.

³¹⁰ H. Frankfurt, ‘The Faintest Passion’, pp. 99 - 100.

³¹¹ R. Stevenson, *The Strange Case of Dr Jekyll and Mr Hyde*, p. 77.

³¹² *Ibid*, p. 80.

Stevenson clearly describes the fundamental changes in the physical characteristics of the body which both individuals share. In her book *Wickedness*, Midgley explains the significance of the physical alteration in the Jekyll/Hyde split as follows: without an understanding of a physical change in either the person of Dr Jekyll or Mr Hyde, we typically have difficulty in understanding how either person could really be seen as radically different persons. Midgley explains:

The disadvantages of oscillating violently in this way are obvious, and in fact if we find people who seem to do it we tend to look for an explanation in some oscillation of their physical state. Without this extra factor, it is hard to see how the oscillator's clarity of vision can really be maintained.³¹³

Initially, once separated, Dr Jekyll no longer associates himself with any of the activities performed by Mr Hyde.³¹⁴ He has become a wholly separate individual and there is no longer a need for him to see any of his Hyde-like qualities as, in any sense, part of his true or real self. This dissociation of Dr Jekyll from Mr Hyde exhibits precisely the kind of externalisation of desires or volitions suggested by Frankfurt and Watson's accounts of self-governance. Frankfurt explains this clearly when he says:

Some of the psychic raw material that we confront may be so objectionable to us that we cannot permit it to determine our attitudes or behaviour. We cannot help having a dark side. However, we are resolved to keep it from producing any direct effect upon the design and conduct of our lives.³¹⁵

What Dr Jekyll begins to realise after his first spontaneous transformation into Mr Hyde, is that despite his dissociation from the latter, he remains inextricably bound to him. I think that it is plausible to say that in realising this, Dr Jekyll glimpses his own self-deception. In his attempt to wholly separate himself from Mr Hyde he thinks, like both Frankfurt and Watson, that he can isolate certain components of his whole self and declare that these aspects no longer truly

³¹³ M. Midgley, *Wickedness*, p. 118.

³¹⁴ "It was Hyde, after all, and Hyde alone, that was guilty". R. Stevenson, *The Strange Case of Dr Jekyll and Mr Hyde*, p. 83.

³¹⁵ H. Frankfurt, 'Taking Ourselves Seriously', p. 10.

belonged to him. I suggest that in practicing this kind of self-deception, Dr Jekyll takes the first step on a path that ultimately leads to the dissolution of his psychological integrity.

After his first spontaneous transformation into Mr Hyde, Dr Jekyll begins to see the danger of attempting to dissociate himself altogether from his Hyde-like qualities. He explains as much, as follows: "I began to spy a danger that, if this [division] were much prolonged, the balance of my nature might be permanently overthrown."³¹⁶ It is from Dr Jekyll's claim that we begin to realise the danger inherent in the process of dissociation. Indeed, such danger is alluded to by psychoanalysts in their discussion of the repression of certain psychic elements – by our censoring and relegating certain desires, or perhaps even other mental elements – to the realm of the unconscious where they can fester and erupt without warning.³¹⁷ At a later point in Stevenson's narrative, Dr Jekyll discovers that Mr Hyde is wanted for the murder of a high-ranking official. Dr Jekyll then makes an attempt at a more total kind of suppression of the part of his nature that he previously transferred into the person of Mr Hyde. The attempt is successful for a short period only, and eventually Jekyll admits the following: "I was still cursed with my duality of purpose; and as the first edge of my penance wore off, the lower side of me [...] so recently chained down, began to growl for license."³¹⁸

Gothic tales such as *Dr Jekyll and Mr Hyde* are meant to shock their readers not only by presenting us with a narrative which forces us to reflect on our own self-division, but by problematising our allegedly natural reactions – our attempts at rejection and dissociation – to our encounters with those seemingly alien aspects of ourselves. The suggestion, then, in much Gothic literature is that by attempting to resolve the internal conflict with rejection and externalisation, persons face the great danger of the dissolution of their psychological integrity and subjectivity.

While there is some truth in Dr Jekyll's statements that, first, to be human is to know conflict between opposing impulses, and, second, that mental life is

³¹⁶ R. Stevenson, *The Strange Case of Dr Jekyll and Mr Hyde*, p. 83.

³¹⁷ See S. Freud, *Introductory Lectures on Psycho-Analysis*.

³¹⁸ R. Stevenson, *The Strange Case of Dr Jekyll and Mr Hyde*, p. 86.

complex and multifaceted, by bringing internal conflict to the forefront, the conflict in Gothic narratives ultimately serve to estrange the characters from their doubles, putting them against one another, which finally results in the characters' desperate attempts to ultimately sever any relation between them. So, in presenting himself as simply an extreme example of "the thorough and primitive duality of man," Dr Jekyll is really being transparently and unsuccessfully, self-deceptive. The consequences, for him, like for all Gothic doubles, are disastrous. The tension between Dr Jekyll's conflicting aspects of himself, and his attempts to dissociate himself from his hidden side (made manifest in Mr Hyde), ultimately results in total disintegration. Such a final acts of total severance or dissociation from the other within results ultimately in the death of the self.

The kind of self-deception at play here is important, and has been described by Midgley as the failure of an agent to acknowledge all the aspects of herself, even those alien aspects, as being constitutive of herself. Midgley suggests that:

self-deception arises because we see motives which are in fact our own as alien to us and refuse to acknowledge them. This is not an isolated event, but is one possible outcome of a very common and pervasive inner-dialogue, in which aspects of the personality appear to exchange views as if they were separate people. We are used to this interchange between alternating moods or viewpoints. (If we were not, we should probably find it much harder to disown some of them, because it would be harder to separate them from our official selves in the first place.)³¹⁹

Gothic fiction illustrates precisely this self-division, as well as the dangers of self-deception, the kind of which Midgley associates with the naive rejection of certain aspects of ourselves as alien or wholly other. While conflict and division are essential features of our experience of our own subjectivity, Midgley argues that it is also an essential feature of our subjectivity that we experience a deep need to bring about inner harmony and integration in our psychic lives. She says:

Integration of the person is not just an optional extra. It is a need. Human beings must have a structure, a policy, a continuity. Each has

³¹⁹ M. Midgley, *Wickedness*, p. 119.

only one life to live. He cannot split up as a coral colony would [...] [C]omplete disintegration, then is hard to imagine. But partial cases are very common. Most of us have personalities fairly well integrated on one side, the side we attend to, but fragmented on others, to which we pay less attention.³²⁰

It is this need which we fail to fulfil when we engage in the kind of self-deception mentioned above. The depiction of the doubles portrayed in Gothic fiction points precisely to the disintegration of psychical unity experienced when practicing Frankfortian externalisation or dissociation, the result of which is a fractured individual who is no longer able to engage in the process of integration.

Furthermore, on Midgley's account, as persons, we manage our internal conflict by trying to achieve a necessary measure of integration and harmony within ourselves. Our efforts at integration and harmony are to a great extent the driving forces behind much of our activity when going about the daily business of living. Who we take ourselves to be, and what we think our character consists of, is constituted by our efforts to achieve a measure of integrity and consistency in the choices we make, the values we endorse and pursue, the commitments and loyalties we stick to, and the goals and projects that we either continue with, or give up on. We also typically think that there must be something wrong with a person, who, as Midgley describes, "drifts from act to act without any attempt at continuity or interest in relating them."³²¹ We would certainly think that such a person's agency would be diminished by being in such a state. Our efforts to maintain a coherent picture of ourselves, then, is expressed in the "deep need which each of us feels to act somehow as a unity," and marks the lifelong project of human personhood. Midgley explains:

When we try, however faintly, to act rather than merely letting forces flow through us, we are not just trying to throw off some outside tyranny. Though there may be such a tyranny, the distinctively free effort surely lies in trying to impose unity on the inner conflict, to decide – as a whole person – what to do. That unity is not given. It is a constantly ongoing project, a difficult, essentially incomplete integration which can occupy our whole lives.³²²

³²⁰ M. Midgley, *Beast and Man*, p. 266-263.

³²¹ *Ibid*, p. 266.

³²² M. Midgley, *The Ethical Primate*, p. 183.

It is at this point, then, that we come to understand Stevenson's warning. His take on duplicity, in its entirety, implies something far graver than the rejection of selective identification and dissociation, purely because they are difficult to achieve. Stevenson's ultimate warning is that when we externalise certain elements of ourselves, as seen in ways recommended by Frankfurt, the result is the loss of the potential to ever be an integrated self. When Stevenson explains that "Utterson knew he was looking on the body of a self-destroyer,"³²³ he gives us far more than a euphemism for the fact that Jekyll/Hyde had committed suicide. Rather, Stevenson expresses the very essence of the moral of his story, in which he "hazard[s] a guess that man will ultimately be known for a mere polarity of multifarious, incongruous and independent denizens."³²⁴ In light of his statements, then, we deny something of fundamental importance about the nature of ourselves when we engage in the process of dissociation. We cannot then, if we take Stevenson's warning seriously, agree that we could achieve self-governance by adopting Frankfurt or Watson's models, which require that we reject or externalise parts of our whole self. The "inner harmony"³²⁵ (the harmony Frankfurt associates with being wholehearted), which would result from this could only be seen as an inner harmony committed to a constant war within the self, and would be akin to a kind of *forcing* ourselves to be the kind of person that we want to be, rather than simply *being* the kind of person we want to be. As Freidman explains:

What ontological importance could there be in the fact (if it is a fact) that the self happens to regard some feature of herself as her "true" self or to "identify" with it? Such identification may be merely a matter of how a person regards herself. Her view that a part of her self is her "true" self, or that her identity does reside in this part rather than that, does not necessarily bring it about that this part is her true self, or is the basis of her identity.³²⁶

If we take the moral of the story presented in *Dr Jekyll and Mr Hyde* to heart, then it requires that we shift our view. Instead of an account of personhood that allows for self-control, in which we understand one part of the self to be more

³²³ R. Stevenson, *The Strange Case of Dr Jekyll and Mr Hyde*, p. 67.

³²⁴ *Ibid.*, p. 76.

³²⁵ H. Frankfurt, 'Taking Ourselves Seriously', p. 17.

³²⁶ M. Friedman, 'Autonomy and the Split-Level Self', p. 29.

definitive than the whole self, we should move to an account that allows for self-governance in which we understand the self as whole. It is only when we take into account the true nature of the whole self that we can properly achieve integration and harmony, which are associated with effective self-governance and personhood.

3.4 Integrity

Following, to some extent, Midgley and Nietzsche, I suggest in this section that the kind of integrity normatively required for personhood requires that we manage our internal conflicts without self-deception or repression. Further, I argue that integrity is required of persons not only within a volitional complex, but over a life.

According to Midgley, persons are to be understood as distinct from non-human animals because they can achieve a kind of integrity which amounts to a wholeness of self. Of non-human animals, she says the following:

crudely speaking, though they do share our struggle to harmonize conflicting motives, they plainly do not have anything like our power of dealing with it by standing back from their various motives, by taking the point of view of the whole, and trying to make some kind of balanced decision.³²⁷

In contrast, the kind of integrity that is unique to a human person involves constant reflection and evaluation of the kind of person we are as a whole – divisions and conflicts included. Through this constant reflection and evaluation, we are able to become clearer about our values, commitments, projects and goals, and so become clearer about our understanding of ourselves as persons. Recall here that, for Midgley, it is through this process of integration that we are responsible for our self-constitution with our whole self in mind, and in keeping our whole-self in mind we exhibit integrity.

Similarly, we have also seen that for Nietzsche self-deception amounts to a lack of integrity in which the self is not considered as a whole self. In the *Genealogy*, Nietzsche says of the noble caste that they were “more whole human

³²⁷ M. Midgley, *The Ethical Primate*, p. 24.

beings.”³²⁸ Furthermore, for Nietzsche, wholeness of self is required for the kind of integrity characteristic of the Sovereign Individual, arguing that if we have achieved integrity:

our ideas, our values, our yeas and nays, our ifs and buts, grow out of us with the necessity with which a tree bears fruit – related and each with an affinity to each, and evidence of *one* will, *one* health, *one* soil, *one* sun.³²⁹

Recall here that on Nietzsche’s account, the antithesis of the Sovereign Individual is characterised by the ‘man of *ressentiment*’. Through an understanding of Nietzsche’s account of *ressentiment*, we can understand his account of integrity as a normative requirement for personhood. Importantly, Frankfurt’s and Watson’s suggestion of a normative requirement for personhood in the form of rejection and dissociation resembles in many ways what Nietzsche envisions as problematic in persons who he sees are guilty of *ressentiment*. The threat here to our personhood, for Nietzsche, is that we practice self-deception when we attempt to solve our internal conflicts by denying that some aspects of our whole self are really an expression of who we truly are. Recall that for Nietzsche, this means, that we might be guilty of *ressentiment*, since genuine affirmation or endorsement of our values or overarching guiding principles requires us to acknowledge, rather than reject or deny, the desires which may conflict with the desire we endorse. Recall further, that this acknowledgment involves the knowledge that the satisfaction of certain desires precludes the fulfilment of other of our desires; and if we do not acknowledge these conflicting desires as our own desires which now cannot reach fulfilment we are in a very important sense deceiving ourselves about ourselves, and are thus guilty of *ressentiment*. This, it seems, is precisely what we would be doing if we followed Frankfurt’s normative prescription for personhood in the form of wholeheartedness.

Frankfurt’s and Watson’s solution to the internal psychological struggle within persons is explained and similarly criticised by Amelie Rorty in terms of a

³²⁸ BG 257.

³²⁹ GM Preface: 2.

practice which she refers to as ‘compartmentalization’ of the self.³³⁰ Those who practice compartmentalisation, according to Rorty, separate and effectively cut off certain aspects of themselves from others. Rorty explains:

Because a divided self tends to undo itself, we naturally, without even being aware of it, attempt to smooth over the appearance of internal conflicts. Or, equally effectively, we compartmentalize attempting to separate different aspects of our lives. Movements toward psychological integration are Janus-faced: they can move toward integrity as well as toward corruption. And although compartmentalization is often effectively soothing, it can rot the mind.³³¹

Importantly, Rorty suggests that the practise of compartmentalisation can be an effective tool for an agent working towards integrity of the self, in which case the compartmentalisation is what she calls ‘soothing’ – by which, I take her to mean that the process is in some sense beneficial to the agent. The practice of compartmentalisation takes a sinister turn, however, when agents reject or deny various compartmentalised aspects of themselves – seeing them as wholly separate from an alleged true or real self. Nietzsche’s understanding of *ressentiment* is in many ways similar here to Rorty’s understanding of the practise of problematic compartmentalisation. It is this more sinister practice of compartmentalisation which lies at the heart of Frankfurt’s accounts of wholeheartedness. Rorty explains the movement towards psychological integration as ‘Janus-faced’, moving either towards corruption or integrity. It is my argument that the more sinister practice of compartmentalisation endorsed by Frankfurt leads to corruption of psychological integrity. For Nietzsche, *ressentiment* is, at least in part, a psychological condition which results in a fractured and damaged individual, lacking in integrity in this sense. The lack of integrity that Nietzsche associates with the ‘man of *ressentiment*’ has to do with the self-deception involved in failing to acknowledge our whole selves. Integrity requires this kind of honesty with ourselves, as Nietzsche explains, it is this honesty which secures integrity. This honesty and integrity is seen in the noble man, a man who he claims is not guilty of *ressentiment*. Nietzsche explains that

³³⁰ A. Rorty, ‘How to Harden Your Heart: Six Easy Ways to become Corrupt’, in Rorty, A., (ed.), *The Many Faces of Evil: Historical Perspectives*, London and New York, Routledge, 2005.

³³¹ *Ibid*, p. 286.

the noble man's "predominance did not lie mainly in physical strength but in strength of soul – they were more whole human beings."³³² In part, however, our integrity is, for Nietzsche, also a reflection of a person's bravery. Nietzsche says that:

While the noble man lives in trust and openness with himself..., the man of *ressentiment* is neither upright nor naive nor honest and straightforward with himself. His soul squints.³³³

There is a kind of bravery seen by Nietzsche in the noble man, which he refers to as 'strength of soul'. This noble bravery is reflected in the Sovereign Individual, who realises that her genuine endorsement of a value involves stifling one or more of her other *own* desires. Such an agent is brave when she owns up to this realisation because she understands that by acting on her endorsed values she is herself both the one who commands herself and the one who obeys. Recall Nietzsche's explanation:

What is called "freedom of the will" is essentially the affect of superiority with respect to something which must obey "I am free 'it' must obey" – this consciousness lies in every will, along with a certain straining of attention, a straight look that fixes on one thing and one thing only, an unconditional evaluation "now this is necessary and nothing else," an inner certainty that it will be obeyed, and whatever else comes with the position of the commander. A person who wills -, commands something inside himself that obeys, or that he believes to obey. But now we notice the strangest thing about the will – about this multifarious thing that people have only one word for. On the one hand, we are, under the circumstances, both the one who commands and the one who obeys, and as the obedient one we are familiar with the feelings of compulsion, force, pressure, resistance, and motion that generally start right after the act of willing.³³⁴

With this realisation the Sovereign Individual bravely exhibits the kind of 'wholeness' of self associated with integrity, while avoiding self-deception.³³⁵ In this respect, then, I take the Nietzschean account to be better suited than Frankfurt or Watson's account to the task of explaining the process of self-governance effectively employed by the person. The person of integrity, for

³³² BG 257.

³³³ GM I: 10.

³³⁴ BG 19.

³³⁵ BG, 257; GM II: 2.

Nietzsche, embraces her whole self, and feels the sting of this 'wholeness' in all of her actions which flow from her evaluations. In feeling this sting, she comes to terms with what Nietzsche would refer to as her 'all-too-human' psychological condition. It is in seeing ourselves as whole selves, and in coming to terms with our psychological condition, that we are most effectively able to exercise our self-governance and personhood. And in aiming to be one's whole-self, though such a self is likely to be fraught with some contradictions, it is my contention that it is, to some extent, these very contradictions which make up who we actually are. Our efforts as persons are not, then, best spent on committing ourselves to a constant war between unsynchronised factions of our selves. Rather, in recognising that all of these factions are components of our 'true' selves, our efforts are best spent in attempting to provide a single, integrated and thus polychromatic montage in which all of these factions are given representation. Even a polychromatic self is naturally subject to certain changes, but is most importantly to be viewed as a whole self, and thus a quest for self-governance can only result from an attempt to maintain the unity of that whole self.

Not unlike Frankfurt and Watson, Nietzsche and Midgley maintain that integrity cannot be achieved without a degree of inner conflict and tension. For the latter, however, the processes of externalisation and dissociation pose a threat to our self-governance because they undermine our integrity as persons. If we see our inner conflicts as essential to who we are as persons, we lack, to some extent, a kind of absolute certainty about ourselves. But this uncertainty is a vital aspect of our subjectivity, and is detrimental to our personhood only insofar as we attempt to disown or disavow it. Moreover, this uncertainty about ourselves is compatible with the kind of integrity that Midgley and Nietzsche see as essential for self-governance. Cox, et al. have summarised this view of integrity in the following convincing way:

The appearance of certainty throughout a person's moral and volitional life may be, and often is, less a sign of integrity than an indication of its lack. Thus, we learn early to beware of the self-righteous and sanctimonious. Such certainty is fostered by different kinds of refusal to reflect upon or acknowledge facts about oneself – a refusal to enter into the fray and nitty-gritty of a life in process. Certainty and self-righteous overconfidence often go hand in hand

with an inexorable consistency of judgment and action. Integrity should be confused with neither.³³⁶

While Frankfurt and Watson's positing of the person of certainty provides one extreme of a lack of integrity, Dr Jekyll provides an example of a person at the other extreme of fragmentation. Cox et al. explain this extreme as follows:

Undue certainty, and concomitant cowardice and self deception, may show a lack of integrity but so does excessive or unwarranted, sometimes feigned, ambivalence. Those who can never make up their minds may be in the grip of a kind of pretence, one reinforced by cowardice and self-deception.³³⁷

Integrity requires that we view ourselves as psychologically predisposed to have internal conflicts. Viewing ourselves as whole selves, acknowledging our internal conflict and tensions, thus promotes integrity and this, we think makes us live well. Integrity, serving as a normative prescription for personhood, should be seen as "a mean between, on the one hand, a disintegrated, utterly fragmented and capricious self and, on the other, a hollow, yet ruthlessly consistent self."³³⁸ Similarly, in support of this view of integrity, Midgley eloquently quotes Jung, as follows:

Painful though it is, this [viewing our darker motivations as part of ourselves] is in itself a gain – for what is inferior or even worthless belongs to me as my shadow and gives me substance and mass. How can I be substantial if I fail to cast a shadow? I must have a dark side also if I am to be whole; and inasmuch as I become conscious of my shadow I also remember that I am a human being like any other.³³⁹

The kind of integrity required for personhood, however, also requires not only integration within the volitional complex, but also over a life. Recall what Midgley says on this point:

³³⁶ D. Cox, M. La Caze, and M. Levine, *Integrity and the Fragile Self*, England, Ashgate Publishing Limited, 2003, p. 3.

³³⁷ *Ibid*, p. 3.

³³⁸ *Ibid*, p. 5.

³³⁹ C. Jung, in M. Midgley, *Wickedness*, p. 126.

Integration of the person is not just an optional extra. It is a need. Human beings must have a structure, a policy, a continuity. Each has only one life to live. He cannot split up as a coral colony would.³⁴⁰

Similarly, Mark Freeman argues that there must be a kind of narrative integrity which unifies a life, saying:

Human lives lived well, I suggest, are anything but the mere succession of moments they are sometimes made out to be or that they can become when they have no point. They are works of Imagination, more unruly than most works of fiction, to be sure, but no less poetic in their connectedness and in the possible beauty of their form.³⁴¹

And similarly, for Nietzsche our lives and indeed our selves bear striking resemblance to works of fiction or biography in that they rely on a similar kind of narrative structure. For Nietzsche “nothing that has happened to us is contingent”³⁴² the occurrence of any given aspect of our lives or selves necessitates all the other aspects of our lives and selves. Nehamas explains the similarity clearly as follows:

[A]bsolutely everything a character does is equally essential to it; characters and the works to which they undoubtedly must belong, are supposed to be constructed so that their every feature supports and is supported by every other one. In the limiting case of the perfect character, no change is possible without corresponding changes in every other feature in order to preserve coherence; and the net result is, necessarily, a different character – taking one part away may always result in the destruction of the whole.³⁴³

Since our lives resemble, for Nietzsche, the lives of literary characters in their structure and composition, the kind of inner harmony or internal coherence with which we associate the Sovereign Individual is a kind of narrative integrity – a coherence of a whole. The kind of narrative integrity seen as a feature of the Sovereign Individual requires the individual to affirm the whole – that is to affirm his existence as a distinctive whole character existing in a particular narrative

³⁴⁰ M. Midgley, *Beast and Man*, pp. 266-263.

³⁴¹ M. Freeman, ‘Death, Narrative Integrity, and the Radical Challenge of Self-Understanding: a Reading of Tolstoy’s *Death of Ivan Ilych*’, *Ageing and Society*, vol. 17, no. 4, 1997, p. 382.

³⁴² A. Nehamas, ‘The Eternal Recurrence’, in Richardson, J., and Leiter, B., (eds.), *Nietzsche*, Oxford, Oxford University Press, 2001, p. 123.

³⁴³ A. Nehamas, ‘How One Becomes What One Is’, in Richardson, J., and Leiter, B., (eds.), *Nietzsche*, Oxford, Oxford University Press, 2001, p. 277.

Further, in his paper 'How One Becomes What One Is', Nehamas suggests that Nietzsche's concern with authenticity is rooted in the fact that much of Nietzsche's thinking revolves around literary models. According to Nehamas, "What is essential to literary characters is their organization"³⁴⁴ – and it is this concern for the organisation of the whole which is central to Nietzsche's characterisation of the Sovereign Individual in the form of his living a good and, I suggest in chapter four, authentic life. In chapter four, I will turn to the question of narrative integrity and authenticity and discuss further the possibilities it opens up for living our lives in ways which would express what Freeman calls 'beauty of form'.

The kind of integrity which I suggest is a normative prescription for personhood requires both that we view ourselves as whole selves over a life, and that we view ourselves as whole selves who are psychologically predisposed to have internal conflicts and tensions throughout that life.

Drawing on Midgley and Nietzsche, then, I have argued that since personhood involves self-governance over and above self-control, personhood normatively requires integrity. In the next chapter, I will argue that there are three further normative requirements for personhood. Persons must meet, I argue, a normative prescription in the form of ambivalence, and an epistemic normative requirement in the form of affirmation, as well as a normative requirement in the form of authenticity.

³⁴⁴ *Ibid*, p. 277.

Chapter 4: Ambivalence, Affirmation, and Authenticity

In the previous chapter, following Frankfurt to some extent, I argued that the concepts of human agency and personhood are to be understood as distinct from one another. I argued that as persons we express ourselves, in part, as individuals through our engagement with the internal conflict required for self-control, and thus, through self-governance. Finally, I argued for a normative prescription for personhood in the form of integrity. In this chapter, I will argue that there is more to being a person than exercising self-governance and living with integrity. As such, I will explore in this chapter three further important normative prescriptions for personhood, all of which centre on the ability to exercise our agency well in the face of our vulnerability.

First, following Nussbaum, I will argue against Frankfurt's normative prescription for personhood in the form of wholehearted wholeheartedness and in favour of a normative prescription in the form of ambivalence, which is its opposite. In his account of personhood Frankfurt endorses a version of value pluralism, arguing that there can be conflicts among our decisive commitments themselves, and that the disparate values underpinning our conflicting commitments will not at all times be commensurable. In *The Fragility of Goodness*, Martha Nussbaum argues in favour of value pluralism on which "valuable things are plural, and are not reducible to some one valuable thing of which all other goods are mere functions."³⁴⁵ Similarly, Lauren Apfel describes pluralism as the belief that there is a "plurality of values," that is, the belief that the many values we encounter in life which "cannot be reduced to or derived from a single master value."³⁴⁶ Pluralists maintain that values are irreducible in that they cannot always be "cashed out into a common currency, ranked into a hierarchy, or related to one another."³⁴⁷ Most importantly, for my purposes in this chapter, pluralists hold that since the valuable things are irreducible, there is always the possibility for insoluble conflict between upholding certain values at

³⁴⁵ M. Nussbaum, *The Fragility of Goodness*, p. xxix.

³⁴⁶ L. Apfel, *The Advent of Pluralism: Diversity and Conflict in the Age of Sophocles*, Oxford, Oxford University Press, 2011, p. 10.

³⁴⁷ *Ibid*, p. 10.

the expense of others. Nowhere is this conflict more evident, for Nussbaum, than in the Greek Tragedies, where the protagonists face “contingent conflicts of value that make it difficult or even impossible for them to pursue all the things to which they have committed themselves.”³⁴⁸ Apfel explains that on the pluralist account:

... because of the ethical complexity of our lives—the reality that many goods will present themselves to us in such a way that their multiplicity cannot be diluted—certain of the values that we cherish and pursue will be unrealizable together. They will be uncombinable in a single human life or society. Some goods will completely preclude others.³⁴⁹

Frankfurt argues in his account of personhood that regardless of the fact that conflict between our decisive commitments can, and does, occur, we ought to cultivate a more global wholeheartedness, or a wholehearted wholeheartedness as an antidote to the possible psychological destruction which results from cases in which we have to sacrifice one of our values for the sake of another. As such, Frankfurt maintains the view, popular in current moral philosophy, that we should be optimistic about the possibility of resolving value conflict.

Against Frankfurt’s view, and following Nussbaum, I will argue in this chapter that “we should be more pessimistic [...] about the possibilities of surmounting [value conflicts, because...] some spheres of value can never be balanced in a way that puts all conflict to rest for all time.”³⁵⁰ Following Nussbaum, I will argue in this chapter that goodness lies not in achieving Frankfurtian wholehearted wholeheartedness (in which values must be ranked and selected between), but in its opposite, ambivalence (in which we recognise that such ranking amongst values is itself problematic). Like Nussbaum, I see conflict between our decisive commitments or values as neither “an obstacle to be overcome,” nor “a problem to be solved,” rather this conflict must be seen as “an irremediable fact of the human condition: not simply a consequence of the scarcity of resources or the brevity of human life, but at times a deep and true reflection of the intrinsic nature of the values themselves.”³⁵¹ Following this I

³⁴⁸ M. Nussbaum, *The Fragility of Goodness*, p. xxix.

³⁴⁹ L. Apfel, *The Advent of Pluralism*, p. 11.

³⁵⁰ M. Nussbaum, *The Fragility of Goodness*, p. xxxi.

³⁵¹ L. Apfel, *The Advent of Pluralism*, pp. 10-11.

will argue in this chapter, contra Frankfurt, for a normative prescription for personhood in the form of ambivalence.

Second, I will argue for a normative prescription for personhood in the form of acknowledgement and affirmation. According to Nussbaum, the ability to recognise the extent to which circumstances, events and other persons play a role “in the planning and conduct of our lives,”³⁵² contributes to our ability to live well, and should this be seen as a normative requirement for personhood. Following Nussbaum, I will argue that it is an epistemic normative requirement for personhood that we recognise our own human fragility and our vulnerability to things over which we have no control. Further, drawing on Nietzsche’s account of the Sovereign Individual, I will argue that to exercise our agency well, or to live well, requires not only that a person recognises their own limits and vulnerabilities, but that they actively affirm their vulnerability. Affirmation here is not to be understood as an emotional state, but rather as an attitude we take to a state of affairs, namely our vulnerability. R. Jay Wallace in *The View from Here: On Affirmation, Attachment, and the Limits of Regret*, defines affirmation along these lines as follows:

To affirm something, in the relevant sense, is to judge that it is valuable along some dimension or other, and also prefer on balance – taking everything into account – that it should not be otherwise than it is.³⁵³

Third, I will argue for a normative prescription for personhood in the form of authenticity. On John Stuart Mill’s account in ‘Of Individuality’, he explores the concept of human authenticity in terms of our ability to direct ourselves to “use and interpret experience in [our] own way”.³⁵⁴ Following Mill, we can understand authenticity as the creative response to our environment which fosters individuality and originality. On such a view, authenticity adds aesthetic

³⁵² M. Nussbaum, ‘Pity and Mercy’, p. 156.

³⁵³ R. Wallace, *The View from Here: On Affirmation, Attachment, and the Limits of Regret*, Oxford, Oxford University Press, 2013, p. 65.

³⁵⁴ J. Mill, *On Liberty*, Harmondsworth, Penguin, 1985, p. 122.

value to our lives, contributing to our living well by making our lives beautiful.

According to Mill:

It is not by wearing down into uniformity all that is individual in themselves, but by cultivating it and calling it forth... that human beings become a noble and beautiful object of contemplation.³⁵⁵

Again, Mill stresses the importance of authenticity in our understanding of ourselves as persons when he says:

Human nature 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.³⁵⁶

Authenticity, then, is inextricably linked to the notion of our individuality and the aesthetic flourish with which it imbues our lives. In this chapter I suggest that it is through being authentic that we might come to practice the ‘art of living’ through which we might come to live most fully as persons.

4.1. Wholehearted Wholeheartedness and Ambivalence

In the previous chapter I raised concerns about Frankfurt’s normative prescription for personhood in the form of wholeheartedness because of its corollary in the form of dissociation and rejection. Given my objections to Frankfurtian wholeheartedness, I suggested that integrity rather than wholeheartedness is a normative prescription for personhood. In this section I raise further concerns about Frankfurt’s normative prescriptions both *for* ‘wholeheartedness’ and *against* ‘ambivalence’. In making his normative prescription of wholeheartedness, not only does Frankfurt suggest that dissociation and rejection are required for wholeheartedness, but he also suggests further that a more global wholehearted wholeheartedness is required to combat the kind of ambivalence which threatens our wholeheartedness on a global scale. Using cases of tragic conflict as a springboard for my discussion, I

³⁵⁵ *Ibid*, p. 127.

³⁵⁶ *Ibid*, p. 123.

argue against Frankfurtian wholehearted wholeheartedness, as it is normatively understood. Following Nussbaum, I argue that it is a normative requirement for personhood that persons ought to experience ambivalence.

Recall that on Frankfurt's view, inner conflicts can arise not only between first and second order desires, but may also arise at the level of our higher order volitions – that is, between and amongst our second order volitions themselves. Recall further that, according to Frankfurt, a person is ambivalent when he experiences a tension within his own 'volitional complex' – the person can be seen to have second order volitions which contradict one another and serve to promote mutually exclusive actions on the part of that person. When a person is ambivalent, then, Frankfurt says that there is "no unequivocal answer to the question of what the person really wants,"³⁵⁷ by which he means that a person is unable to make a decisive commitment to one of his second order volitions. The ambivalent person is volitionally conflicted and thus, despite having engaged in the process of critical reflection on her first order desires, such a person has no one particular point of view from which to act which could be seen as her point of view.

Following this, Frankfurt suggests a normative understanding of ambivalence in which ambivalence threatens our personhood, because it undermines our ability to be particular subjects – that is, to have made decisive commitments about our second order volitions, and thus to have formed a particular point of view from which to speak which can be seen as our own particular, subjective, point of view. On Frankfurt's view, the ambivalent person vacillates between volitions, and this manifests in their lives as inconstancy in action, and the inability to form a coherent picture of themselves as subjects precisely because of a lack of decisiveness about their volitions. For Frankfurt, ambivalence is thus to be avoided, normatively speaking, because it has psychological consequences which limit rather than promote our personhood. Accordingly, Frankfurt explains the pathology underlying the classical Freudian case of the Rat Man – one of Freud's most famous case studies – as a case of ambivalence. Frankfurt argues that in the Rat Man's case, his volitional

³⁵⁷ H. Frankfurt, 'Identification and Wholeheartedness', p. 165.

vacillation between love and hatred for his father could not be overcome by self-deception or by repression, but can be solved only by his decisive commitment to one or another of his volitions. In Frankfurt's words:

In order for the Rat Man to be wholeheartedly on the side of his benign attitudes, it would not have been necessary for him to conceal his hostile feelings from himself. Nor would he have had to refrain from making a conscious effort to deal with those feelings in whatever ways might be effective and helpful. It would have required only that, in the struggle between his hatred and his love for his father, he himself come to stand decisively against the hatred and behind the love.³⁵⁸

Frankfurt explains again as follows:

Surely it is ambivalence, and not wholeheartedness, that is a disease of the will...the Rat Man's symptoms "often involved repeatedly doing and undoing an action, or thinking and contradicting a thought." That sort of self-defeating behavior and thought violates the elementary requirements of rationality. It is not a consequence of repression as such, but ambivalence. It is a manifestation of the incoherence in which, precisely, the divided will of ambivalence consists. The desire for wholeheartedness is nothing other than a desire to be free of this crippling irrationality.³⁵⁹

Through making a decisive commitment, we become, for Frankfurt, wholehearted in our commitments, and it is through the psychic unity that wholeheartedness brings, that our personhood is promoted. Thus, being wholehearted, according to Frankfurt, is the cure to the Rat Man's ambivalence, and indeed to all other cases of ambivalence of this sort.

In *Necessity, Volition and Love*,³⁶⁰ published over a quarter of a century after 'Freedom of the Will and the Concept of a Person', Frankfurt explicitly develops his account of personhood in which the notions of ambivalence and wholeheartedness play a central role. Here Frankfurt suggests that over and above the self-control required for agency, personhood requires that we live in accordance with the ideals which we have set for ourselves through our decisive commitments. Wholeheartedness requires, for Frankfurt, that we stand

³⁵⁸ H. Frankfurt, 'Reply to J. David Velleman', in Buss, S., and Overton, L., (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt*, United States of America, MIT Press, 2002, p. 126.

³⁵⁹ *Ibid*, p. 127.

³⁶⁰ H. Frankfurt, *The Importance of What we Care About*.

resolutely behind our decisive commitments, and that our decisive commitments give us normative reasons for action – that is, when we fail to act in accordance with our decisive commitments we fail to exercise our agency well. Importantly, Frankfurt argues that while our agency is undermined by actions which are compulsive or compelled, as persons we are subject to further normative constraints on our actions. He calls actions which are neither compelled nor compulsive ‘wholly voluntary’ in the sense that our agency has not been undermined (we have still exercised self-control in such cases), but argues that, as persons, there are circumstances in which we have ‘no real alternatives’ to performing certain of these voluntary actions, normatively speaking. According to Frankfurt, our being wholehearted requires that we perform certain voluntary actions *necessarily*. That is, in making our decisive commitments, we are now normatively constrained by the things which we have made decisive commitments about.

As persons, Frankfurt plausibly argues, we are constrained in our actions by the goals and values we set for ourselves through our decisive commitments in three distinct ways: by what Frankfurt calls the ‘*necessities of ambition and prudence*’; the ‘*necessities of duty*’; and the ‘*necessities of love*’. Frankfurt suggests that we are normatively bound in our actions by the commitments we have made to achieve our settled goals, we are normatively bound by a set of moral obligations, and we are bound by the commitments we make to the things and persons we genuinely care about. Importantly, according to Frankfurt, if we fail to act in accordance with our decisive commitments – that is, when we fail to act from the necessity given by our decisive commitments – we fail to be wholehearted. Any failure in wholeheartedness is described by Frankfurt as a form of self betrayal which results in ‘drastic psychic injuries’ to the person, which compromises our ability to be good agents, or to live well. He says:

A person who fails to act in the ways that caring about his beloved requires necessarily fails to live in accordance with his ideal for himself. In betraying the object of his love, he therefore betrays himself as well. Now the fact that a person betrays himself entails of course, a rupture in his inner cohesion or unity; it means that there is a division within his will. There is, I believe, a quite primitive human need to establish and to maintain volitional unity. Any threat to this unity – that is, any threat to the cohesion of the self – tends to alarm a person and to mobilize him for

an attempt at “self-preservation.” ... He is opposing ends and interests that are essential to his nature as a person. In other words, he is betraying himself. We are naturally averse to inflicting upon ourselves such drastic psychic injuries.³⁶¹

An important question about wholeheartedness should be raised at this point in Frankfurt’s account: what about the possibility of conflict between our decisive commitments themselves? Frankfurt argues, I think importantly and problematically that in those rare instances in which a person is torn between two opposing decisive commitments which are ‘equally essential to their nature as a person’, the self-betrayal involved in acting on either one of those decisive commitments undermines their personhood in a fundamental way. Since personhood involves for Frankfurt expressing ourselves as a particular individual, when our personhood is undermined by self-betrayal, we are no longer that same particular individual who we were before. In Frankfurt’s words:

Situations in which it is impossible for a person to avoid this sort of self-betrayal provide the theme for one variety of human tragedy. Thus, Agamemnon at Aulis is destroyed by an inescapable conflict between two equally defining elements of his own nature: his love for his daughter and his love for the army he commands. His ideals for himself include both being a devoted father and being devoted to the welfare of his men. When he is forced to sacrifice one of these, he is thereby forced to betray himself. Rarely, if ever, do tragedies of this sort have sequels. Since the volitional unity of the tragic hero has been irreparably ruptured, there is a sense in which the person he had been no longer exists. Hence, there can be no continuation of *his* story.³⁶²

Frankfurt suggests, then, that in cases of tragic conflict in which a person is truly ambivalent with regard to having to act on only one of two competing decisive commitments (that is, in cases where an agent must, when acting, sacrifice one genuine value for the sake of another genuine value), an agent’s personhood is undermined in that they are no longer the same person they were before, after acting. Similarly Velleman says of Frankfurt’s account:

³⁶¹ H. Frankfurt, ‘Autonomy, Necessity, and Love’. in Frankfurt, H., *Necessity, Volition and Love*, Cambridge, Cambridge University Press, 1999, p. 139.

³⁶² *Ibid*, Footnote 8, p. 139.

[A] threat to the self, as conceived by Frankfurt, comes from ambivalence, which guarantees that one or another element of the self will have to be sacrificed when a choice is made. If the self is to have any chance of remaining whole, it must be wholehearted, in the sense of being unequivocal in its essential concerns.³⁶³

If we care about our selves as persons, it seems to follow normatively from Frankfurt's account that we ought to strive to avoid cases of ambivalence between our decisive commitments. We can do this, Frankfurt suggests, by cultivating a more global wholeheartedness amongst our decisive commitments, or a wholehearted wholeheartedness we might call it.

In the case of the Rat Man discussed above, being wholeheartedly wholehearted requires that the Rat Man make a decisive commitment to resolve his ambivalence. Cases of tragic conflict, however, provide us with the most extreme and interesting cases of conflict at the level of our decisive commitments. In what follows I will argue contra Frankfurt that that in cases of tragic conflict we ought not, like the Rat Man, to make a decisive commitment to resolving our ambivalence between our decisive commitments.

In his discussion above, Frankfurt uses the ancient Greek tragedy of Agamemnon as an illustrative example of a case of tragic conflict in which there a person is ambivalent between their decisive commitments. What makes the case of Agamemnon of particular interest is that, at least according to Frankfurt's characterization of Agamemnon's situation, he is caught in a situation in which whatever he does, he will commit an act of self-betrayal and his personhood will be undermined. Accordingly, for Frankfurt the tragedy of Agamemnon's case lies precisely in the fact that he is obligated by two competing and contradictory, yet both normatively required, actions as determined by his own decisive commitments. Unlike the Rat Man then, his ambivalence cannot in principle be resolved because there is no way for the person to be 'unequivocal' in his final decisive commitment in his situation, and there is no way for him to resolve this conflict. Since the self-betrayal involved in Agamemnon's actions undermines his personhood, it follows that on Frankfurt's account it would be all things considered better if Agamemnon were to be able to wholeheartedly commit

³⁶³ J. Velleman, 'Identification and Identity', in S. Buss, and L. Overton, (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt*, United States of America, MIT Press, 2002, p. 95.

himself to either of his competing decisive commitments. Nussbaum, however, provides us with a contrary, and I think convincing, reading of the tragedy in Agamemnon's case which should lead us to reject Frankfurt's suggestion that self-betrayal and ambivalence undermine personhood by undermining our ability to live well. Further, following Nussbaum, I argue that personhood normatively requires that person's ought to experience ambivalence, and ought not to resolve conflicts between their decisive commitments through becoming wholeheartedly wholehearted.

In his characterization of Agamemnon's case, Frankfurt supports the claim that a commitment to value pluralism entails that there can be cases of tragic conflict in which persons are forced to betray themselves in their actions. For the value monist, even in Agamemnon's dilemma, there is only one right course of action to take. If Agamemnon were to take this 'right' course of action, there would be no reason for him not to be wholeheartedly wholehearted about his choice, and so his lack of wholehearted wholeheartedness would amount to a failing on his part to selectively identify with the one, clearly right choice. Nussbaum argues, here, that for value monists, in the case of Agamemnon's dilemma, "at most one requirement could be legitimate or valid, if we find it, it is the true obligation, and the other one naturally ceases to exert any claim on our attention."³⁶⁴ But as Frankfurt suggests, in the case of Agamemnon, there is no clearly right choice to make between his two competing volitions. And it is precisely in this incommensurability of values for Agamemnon that leads Frankfurt to his conclusion that tragic conflict necessarily leads to self-betrayal – that is, the kind of self betrayal involved in acting against our own decisive commitments or genuine values.

In *The Fragility of Goodness*³⁶⁵, Nussbaum argues that a commitment to strong value pluralism entails the possibility of a person finding themselves in a situation where a conflict between their values cannot be resolved, or in which it is made impossible for a person to be wholeheartedly wholehearted. Nussbaum says:

³⁶⁴ M. Nussbaum, *The Fragility of Goodness*, p. 36.

³⁶⁵ M. Nussbaum, *The Fragility of Goodness*.

The recognition of a plurality of genuine goods always leaves open the possibility of conflict; thus we should be more pessimistic than Hegel was about the possibilities of surmounting them. I continue to believe that this is basically correct: some spheres of value can never be balanced in a way that puts all conflict to rest for all time.³⁶⁶

Value pluralism thus entails, for Nussbaum, that persons will, from time to time, be subject to “contingent conflicts of value that make it difficult or even impossible for them to pursue all the things to which they have committed themselves.”³⁶⁷ According to Nussbaum these situations expose the ways in which persons are “vulnerable to luck,”³⁶⁸ and she argues that we can distinguish between cases in which luck affects not merely our ‘contentment, but ethical goodness itself’. She explains as follows:

We are considering situations, then, in which a person must choose to do (have) either one thing or another. Because of the way the world has arranged things, he or she cannot do (have) both, or, regardless of what he actually wants, he has some reason to do (have) both. Both alternatives make serious claim upon his practical attention. He senses that no matter how he chooses he will be left with some regret that he did not do the other thing. Sometimes the decision itself may be difficult; his concerns seem evenly balanced. Sometimes he may be clear about which is the better choice, and yet feel pain over the frustration of other significant concerns... We want ultimately to ask whether among these cases there are some in which not just contentment, but ethical goodness itself, is affected: whether there is sometimes not just the loss of something desired, but actual blameworthy wrongdoing – and therefore, occasion not only for regret but for an emotion more like remorse.³⁶⁹

Agamemnon, she argues, provides us with an example in which ethical goodness itself is at stake. On Nussbaum’s detailed reading, Agamemnon commits an action “without any direct physical compulsion and in full knowledge of its nature,”³⁷⁰ that is to say, he exercises the self-control central to human agency, but is constrained by his situation which forces him to choose between two equally important values which he selectively identifies with. Agamemnon is forced by his situation to commit an action about which it can be said that his “ethical

³⁶⁶ *Ibid*, p. xxxi.

³⁶⁷ *Ibid*, p. xxix.

³⁶⁸ *Ibid*, p. xxix.

³⁶⁹ *Ibid*, p. 27.

³⁷⁰ *Ibid*, p. 25.

character or commitments would otherwise dispose him to reject the act.”³⁷¹

Nussbaum explains:

Agamemnon is allowed to choose: that is to say, he knows what he is doing; he is neither ignorant of the situation nor physically compelled; nothing forces him to choose one course rather than the other. But he is under necessity in that his alternatives include no very desirable options.³⁷²

In Aeschylus’s tragedy, Agamemnon explains his dilemma as one in which he will be forced, whichever way he acts, to betray his own goals and values set by his decisive commitments. Agamemnon says:

“A heavy doom is disobedience, but heavy too, if I should rend my own child, the adornment of my house, polluting a father’s hands with streams of slaughtered maiden’s blood close by the altar. Which of these is without evils?”³⁷³

Importantly, however, Nussbaum argues that even in Agamemnon’s case, there is still room for ethical assessment of Agamemnon in his situation, and this assessment has everything to do with his wholehearted wholeheartedness. According to Nussbaum’s reading the case of Agamemnon differs from the biblical tale of Abraham in an important way, namely in difference between Abraham and Agamemnon’s reactions to their ambivalence. While initially similar, in that both Agamemnon and Abraham find themselves in situations in which generally good men must “either kill an innocent child out of obedience to a divine command, or incur the heavier guilt of disobedience and impiety,”³⁷⁴ the two tragic tales, she argues, diverge radically from this point onwards. According to Nussbaum, in Abraham’s case, we see a “delicate struggle between love and pious obligation [...] followed by a sacrifice executed with horror and reluctance.”³⁷⁵ Abraham, it seems, remains ambivalent about his actions, torn between two decisive commitments which are equally constitutive of himself, even when the time came to act. Though Abraham is not made impotent in his

³⁷¹ *Ibid.*, p. 25.

³⁷² *Ibid.*, p. 34.

³⁷³ Aeschylus in *Ibid.*

³⁷⁴ *Ibid.*, p. 35.

³⁷⁵ *Ibid.*

ability to settle on and choose a course of action - choosing to act in accordance with the necessities of duty – it remains true of him that he remains ambivalent about the choice he has made. Nussbaum claims that Agamemnon’s tale is quite different to the tale of Abraham, precisely because Agamemnon should be seen as wholeheartedly embracing his choice and overcoming his ambivalence. Nussbaum explains this wholehearted wholeheartedness as an “unnatural cooperation of internal with external forces,”³⁷⁶ and claims that unlike Abraham who remains conflicted about his choice of action and continues to rile against the hand that fate has dealt him, Agamemnon accepts his situation and has no remorse or regret about executing his chosen course of action. In wholeheartedly embracing one rather than another of his decisive commitments, Agamemnon, unlike Abraham, turns himself into a ‘collaborator’ with fate and a ‘willing victim’ of circumstances. Nussbaum explains:

Agamemnon now begins to cooperate inwardly with necessity, arranging his feelings to accord with his fortune. From the moment he makes his decision, itself the best he could have made, he strangely turns himself into a collaborator, a willing victim.³⁷⁷

Given his situation, Nussbaum claims that we would expect Agamemnon, who has been characterised as a good king and loving father in the narrative thus far, to react with ‘pain and revulsion’ to having settled on a course of action in which he would sacrifice his own daughter, and betray his decisive commitment to being a good father. How Agamemnon in fact approaches his choice, Nussbaum argues, radically undermines this basic ethical expectation, and challenges his ‘goodness itself’. According to Nussbaum’s translation Agamemnon reacts to his decision with wholehearted acceptance, saying:

“For it is right and holy that I should desire with exceedingly impassioned passion... the sacrifice staying the winds, the maiden’s blood. May all turn out well.”³⁷⁸

³⁷⁶ *Ibid.*

³⁷⁷ *Ibid.*

³⁷⁸ Aeschylus in *Ibid.*

Nussbaum argues that there are two important things we should take note of in light of what Agamemnon says here:

Agamemnon seems to have assumed, first, that if he decided right, the action chosen must be right; and, second, that if an action is right, it is appropriate to want it, even to be enthusiastic about it.³⁷⁹

Nussbaum argues that Agamemnon's attitude toward his chosen course of action "seems to have changed with the making of it".³⁸⁰ While his description of his situation made earlier in the narrative shows his reluctance to perform either of his possible actions, once he has made the choice between alternatives, Nussbaum argues that he now takes an attitude of "peculiar optimism,"³⁸¹ saying:

An act we were prepared to view as the lesser of two hideous wrongs and impieties has now become for him pious and right, as though by some art of decision-making he had resolved the conflict and disposed of the other 'heavy doom'.³⁸²

Along with his new attitude towards his decision Agamemnon also sees his choice as justified, since he "has chosen the *better* course, all may yet turn out *well*."³⁸³ Further, and importantly, in Agamemnon seeing his choice as the better or right choice, he also comes to think that it is good for him to be wholeheartedly wholehearted in his choice and embrace it with vigour. Nussbaum says:

At the same time, we notice that the correctness of his decision is taken by him to justify not only action, but also passion: if it is right to obey the god, it is right to *want* to obey him, to have an appetite for the crime, even to yearn for it with impassioned passion.³⁸⁴

Following Nussbaum, then, *contra* Frankfurt there can be a sequel to Agamemnon's story because Agamemnon overcomes his ambivalence through wholehearted wholeheartedness, and thus, ultimately, Agamemnon does not

³⁷⁹ *Ibid*, pp. 35-36.

³⁸⁰ *Ibid*.

³⁸¹ *Ibid*.

³⁸² *Ibid*.

³⁸³ *Ibid*.

³⁸⁴ *Ibid*.

seem to be guilty of the Frankfurtian sort of self-betrayal – he is in fact wholeheartedly wholehearted about his decisive commitments. But Agamemnon seems to me to be guilty of a deeper kind of self-betrayal, and, according to Nussbaum, the tragedy in Agamemnon’s case is the result of what she sees as a normative transgression on his part.

Nussbaum claims that the inescapable dilemma in which Agamemnon finds himself – either failing in his duty as King and divine subject or sacrificing his own daughter – results from purely “external and contingent”³⁸⁵ factors and not from any “personal guilt of Agamemnon’s.”³⁸⁶ The tragic conflict is seen in the *Oresteia*, according to Nussbaum, because Agamemnon commits the transgression willingly, and indeed wholeheartedly, and can thus be subject to moral assessment - despite the fact that we think Agamemnon would, given a different set of alternatives, perhaps not commit the act at all. In support of her argument in favour of finding moral fault with Agamemnon’s actions she draws on textual support from the lines of the Chorus, saying:

The Chorus does not so much blame the fact of the action, for which they feel the gods bear a primary responsibility... What they impute to Agamemnon himself is a change of thought and passion accompanying the killing for which they clearly hold him responsible. He put up with it. He did not struggle against it... Never, in the choral narration or subsequently, do we hear the king utter a word of regret or painful memory.³⁸⁷

Similarly, she draws on the Chorus in the tragedy of Eteocles, to argue in support of her moral assessment of Agamemnon, saying again:

But what we, with the women of the chorus, feel most clearly is, as in Agamemnon’s case, the perversity of the king’s imaginative and emotional response to this serious practical dilemma. He appears to feel no opposing claim, no pull, no reluctance. He goes ahead with eagerness, even passion. It is around these deficiencies of vision and response that the blame of the chorus centres: his eagerness, his bad eros, his bestially hungry desire. Whether or not they would have chosen differently, they are clear that he has made things too simple. He has failed to see and

³⁸⁵ *Ibid*, p. 34.

³⁸⁶ *Ibid*, p. 33.

³⁸⁷ *Ibid*, pp. 36-37.

respond to his conflict as the conflict it is, this crime compounds the already serious burden of his action.³⁸⁸

What makes Agamemnon, and equally Eteocles, morally blameworthy in Nussbaum's opinion, is both that he fails to recognise "the situation as a situation that forces him to act against his character," and that he fails to acknowledge that this is a case in which he "will have to do wrong" because the case is not soluble.³⁸⁹ It is, for Nussbaum then, the recognition on our own part of a genuine plurality of goods which is essential to our moral character, further, in recognising this, the person of good moral character will cultivate a kind of ambivalence in which he still feels "the force of the losing claim", and in Agamemnon's case "the demand of good character for remorse and acknowledgement."³⁹⁰ Importantly, it is through ambivalence that the person of good moral character will "feel and exhibit the feelings appropriate to a person of good character caught in such a situation," Nussbaum says:

He would not regard the fact of decision as licensing feelings of self-congratulation much less feelings of unqualified enthusiasm for the act chosen. He will show in his emotive behaviour, and also genuinely feel, that this is an act deeply repellent to him and to his character. Though he *must*, to some extent, act like a person 'who is called by the worst names', he will remember, regret and, where possible make reparations. This emotion, moreover, will not be simply regret, which could be felt and expressed by an uninvolved spectator and does not imply that he himself has acted badly. It will be an emotion more like remorse, closely bound up with acknowledgement of the wrong that he has as an agent, however reluctantly, done.³⁹¹

Following Nussbaum's argument, then, it would seem that it is ambivalence, rather than wholehearted wholeheartedness, which ought to serve as a positive normative requirement for personhood. To this effect, Nussbaum argues that is 'human goodness itself' which is at stake in our pursuit of either wholehearted wholeheartedness or ambivalence, saying in favour of ambivalence:

If we were such that we could in a crisis dissociate ourselves from one commitment because it clashed with another, we would be less good.

³⁸⁸ *Ibid*, p. 39.

³⁸⁹ *Ibid*, p. 42.

³⁹⁰ *Ibid*, p. 49.

³⁹¹ *Ibid*, p. 43.

Goodness itself, then, insists that there should be no further or more revisionary solving.³⁹²

I think that Nussbaum's suggestion here is intuitively plausible. If we accept value pluralism, being able to make a decisive commitment in the face of a clash between two of our genuine commitments would, *contra* Frankfurt, look quite clearly to us like an act of self-betrayal.

This kind of self-betrayal is often associated in popular discourse with what are sometimes called 'Sophie's choice cases' originating from the case of Sophie presented in William Styron's novel *Sophie's Choice*.³⁹³ In that novel, Styron describes the case of Sophie, a woman detained with her two children in a Nazi concentration camp during the Second World War. Sophie is forced to make a choice between which one of her children will live and which one will die by a Nazi doctor:

... the doctor said, "You may keep one of your children"
"Bitte?" said Sophie.
"You may keep one of your children," he repeated. "The other one will have to go. Which will you keep?"
"You mean I have to choose?"
... Her thought process dwindled, ceased. Then she felt her legs crumple. "I can't choose! I can't choose!" She began to scream...
Tormented angels never screeched so loudly above hell's pandemonium.
"Ich kann nicht wahlen!" she screamed.
The doctor was aware of the unwanted attention. "Shut up!" he ordered.
"Hurry now and choose. Choose goddamnit, or I'll send them both over there. Quick!"
.. "Don't make me choose," she heard herself plead in a whisper, "I can't choose."
"Send them both over there, then," The doctor said to the aide, "nach links."
"Mama!" She heard Eva's thin but soaring cry at the instant that she thrust the child away from her and rose from the concrete with a clumsy stumbling motion. "Take the baby!" she cried out. "Take my little girl!"³⁹⁴

Cases of tragic conflict, like Sophie's or Agamemnon's, expose in my opinion the shortcomings of Kantian and Utilitarian approaches to ethics, as well as so called 'could-have-done-otherwise' approaches to moral responsibility. On all of these approaches we can make a moral assessment of Sophie's choice. As I see it, such

³⁹² *Ibid*, p. 50.

³⁹³ W. Styron, *Sophie's Choice*, New York, Random House, 1976.

³⁹⁴ *Ibid*, pp. 483-484.

moral evaluation is misguided. Sophie's choice, it seems to me, is less important than the way Sophie goes about choosing – that is, what is most important here is how Sophie has made the choice, or how Sophie has approached and reacted both in and to her situation. Similarly, if ambivalence is a normative requirement for personhood, our moral evaluation is directed at Sophie herself, asking essentially how we judge Sophie's character. In Styron's novel, Sophie reacts to her situation in precisely the opposite way to Agamemnon, in Aeschylus's play. If we agree with Nussbaum and if we accept ambivalence as a normative prescription for personhood, then we should have a favourable moral judgment of Sophie regardless of what her actual choice was precisely because our moral judgment should be directed to the way in which Sophie responds to her situation and her choice.

If this is the case, we might expect a happy ending for Sophie's case, since she has, after all, acted as the virtuous person would in her circumstances – if Sophie did the right thing by being ambivalent, unlike for Agamemnon, all “may yet turn out well”. But Styron's novel is no less tragic than Aeschylus's play, and Sophie's eventual suicide is the culmination of a life we would be hard pressed to describe as happy. Part of the tragic story in Sophie's case is the effect of Sophie's choice on her life following her choosing. Sophie's ambivalence becomes all consuming, and she is never able to overcome her loss, or her grief and guilt. In light of this, it might be argued in favour of Frankfurt's normative prescription for personhood in the form of wholeheartedness, that were Sophie to have resolved her ambivalence through her choice, she might have been able to better come to terms with that choice, and she might have been able to, after a suitable period of mourning and regret, go on living out the rest of her life in this world, opening herself up to new experiences and what joy the world may still have to offer her.

I think there is little evidence to suggest that if Sophie were to have acted wholeheartedly she would have been able, eventually, to enjoy the rest of her life, but her life would nevertheless have been compromised from the ethical point of view. Having acted wholeheartedly in her situation, if it could have

been at all possible,³⁹⁵ would have been an act of self-betrayal on her part, which I think she would similarly be hard pressed to 'get over'. Her genuine love for two incommensurable goods, her children, precludes her acting wholeheartedly wholehearted without deeply rupturing her sense of self which would have caused her to have emotionally similar reactions to the case in which she exhibits unwavering ambivalence. It seems clear to me, then, that in tragic cases, such as those of Agamemnon and Sophie, ethics and happiness may well have to come apart. Aristotle himself thought that some lives, however virtuously lived, can nevertheless, in light of their tragic circumstances, fail to be flourishing lives. Horsthouse explains this intuitively plausible Aristotelian insight as follows:

'Good action' is so called advisedly, and although it is conceptually linked to morally correct (right) decision and to 'action of the virtuous agent', it is also conceptually linked to 'good life' and *eudiamonia*.

The actions a virtuous agent is forced to in tragic dilemmas fail to be good actions because of the doing of them, no matter how willingly or involuntary, mars or ruins a good life. So to say that there are some dilemmas from which even a virtuous agent cannot emerge having acted well is just to say that there are some from which even a virtuous agent cannot emerge with her life unmarred...³⁹⁶

Even more importantly, the deeper tragedy at work in *Sophie's Choice* stems from our acknowledgement of the tension between Sophie's acting well (by being ambivalent) and Sophie's tragic life. Here is revealed most poignantly the tragic truth of life: that a life lived well does not necessitate happiness. And in this revelation, we rile at the tragic story because of our distaste for and uncomfortableness with the deep injustice of our human situation. But this should not overshadow a crucial aspect of Sophie's story which distinguishes it from that of Agamemnon's, and which is of greatest significance to my claim that ambivalence is a normative prescription for personhood. Our judgment of Sophie's character is favourable because she acknowledges the full weight of both her situation and her actions, and it is unclear whether our judgment of her would remain favourable if her life story did not reflect this tragedy.

³⁹⁵ I think that Sophie's case presents us with a situation in which wholeheartedness might genuinely not be possible. Sophie loves both of her children, and it would be difficult to imagine that she could be wholehearted about a decision to have either one of them killed.

³⁹⁶ R. Hursthouse, *On Virtue Ethics*, New York, Oxford University Press Inc., 1999, p. 74.

Cases of tragic conflict, then, help us to understand what is normatively required of persons, that is, to exercise our agency well. What is required, normatively of persons, is that they experience ambivalence in cases of tragic conflict. Of course, we are hopeful that we may never ourselves end up in situations as desperate as those dreamed up by the tragedians, nevertheless the normative insight highlighted through tragedy is no less important for us as we go about our daily lives. If we accept value pluralism, we will face real and significant conflicts between our values from time to time, perhaps even most of the time, and this makes us, as persons, extremely vulnerable sorts of beings. I may have to choose to sacrifice familial values for the sake of my professional values, or I may have to sacrifice social values for the sake of my political values. In many of these cases, my choices to sacrifice something of value should be accompanied by a sense of regret, and sometimes remorse. Though I may ultimately be satisfied by my choices, in remaining ambivalent about them, I pay homage to both values, even when only upholding one in my actions. It is through our ambivalence in the face of a plurality of goods that we are able to truly understand ourselves as the kinds of creatures that we are, that is, creatures living complex and dynamic lives where values often place competing, even tragic, demands on us. Persons are vulnerable beings because the world in which they live.

In this section I have argued, contra Frankfurt, for a normative prescription for personhood in the form of ambivalence. Following Nussbaum I argued that cases of tragic conflict highlight the importance of ambivalence for exercising our agency well in the face of our vulnerability. In the next section, following Nussbaum's lead again, at least to some extent, and drawing on Nietzsche's account of the Sovereign Individual, I turn my attention to an epistemic normative prescription for personhood. Persons, I argue, ought to not only to acknowledge, but affirm, their vulnerability.

4.2. Acknowledgment and Affirmation

In this section I will argue further against Frankfurt's characterisation of personhood, and argue for an epistemic normative prescription for personhood in the form of our own acknowledgement and affirmation of our vulnerability. To exercise our agency well we cannot, I argue, attempt to transcend our vulnerability as Frankfurt, like the Stoics, seems to suggest. In her paper 'Pity and Mercy: Nietzsche's Stoicism' Martha Nussbaum argues that Nietzsche's philosophical project can be seen as an attempt to "bring about a revival of Stoic values of self-command and self-formation."³⁹⁷ She argues that, to his detriment, Nietzsche's Sovereign Individual epitomises a kind of Stoic ideal of inner strength and self-sufficiency which "goes beyond Stoicism" in its valorisation of radical self emancipation from the contingencies of life and from our own human vulnerability. Nussbaum thus urges us to question whether the picture of strength in Nietzsche's Sovereign Individual is really a picture of human strength at which we would be willing to, or at which we ought to, aim. In this section I challenge Nussbaum's criticism of the Sovereign Individual arguing that Nietzsche and Nussbaum have more in common than she suggests. I argue, following both Nietzsche and Nussbaum, that it is an epistemic normative requirement for personhood that we acknowledge and affirm our own vulnerability.

One important reason we value human agency, is that it is through our agency we have the ability to make our lives less subject to the contingencies of living in a world which is largely out of our control. Human agency, which I have explained as our unique ability to guide our selves and lives in a world which is indifferent to our desires and efforts, relies in part on our ability to increase our control over our internal situation. That is, in a world in which we largely cannot control external circumstances, we are, perhaps uniquely, situated by our ability to control our internal psychological conditions and hopefully the actions and behaviours which flow from them. On such an account, as I have argued so far in this thesis, what is fundamental is self-control, or control over our internal

³⁹⁷ M. Nussbaum, 'Pity and Mercy', p. 140.

psychological situation. This picture is clearly illustrated by Frankfurt when he explains:

For to deprive someone of his freedom of action is not necessarily to undermine the freedom of his will. When an agent is aware that there are certain things he is not free to do, this doubtless affects his desires and limits the range of choices he can make. But suppose that someone... has in fact lost or been deprived of his freedom of action. Even though he is no longer free to do what he wants to do, his will may remain as free as it was before. Despite the fact that he is not free to translate his desires into actions or to act according to the determinations of his will, he may still form those desires and make those determinations as freely as if his freedom of action had not been impaired.³⁹⁸

For Frankfurt, whether or not our actions are in fact limited by our situation, or indeed when our freedom of action has been entirely constrained, the freedom which is available to all human agents, in all circumstances, cannot be undermined in this way because, as Frankfurt puts it, "he may still form those desires and make those determinations as freely as if his freedom of action had not been impaired" - though we may not all actually exercise this freedom at any given time, or ever. The most persuasive cases for the centrality of specifically self-control as central to our understanding of agency are made by appealing to our intuitions about what happens to agents in situations of extreme constraint, as Frankfurt says cases in which someone "has in fact lost or been deprived of... freedom of action". Recall here Viktor Frankl's discussion in *Man's Search for Meaning*, in which a kind of inner freedom or self-control remains available to us in even the most restrictive and oppressive circumstances such as a concentration camp. According to Frankl, at the very last we have control over our internal mental and psychological states, and this is the kind of control which external circumstances cannot have an effect on. He writes:

Every day, every hour, offered the opportunity to make a decision, a decision which determined whether you would or would not submit to those powers which threatened to rob you of your very self, your inner freedom; which determined whether or not you would become the plaything of circumstance, renouncing freedom and dignity to become moulded into the form of the typical inmate.³⁹⁹

³⁹⁸ H. Frankfurt, 'Freedom of the Will and the Concept of a Person', pp. 14-15.

³⁹⁹ V. Frankl, *Man's Search for Meaning*, p. 74.

On such a view, when I am faced with a world in which the ends and goals I have conceived of are made unattainable, I need not necessarily feel my agency restricted or diminished because my agency is constituted by my self-control – as Frankfurt would put it, we still have freedom of the will. Nothing and no-one outside of me can truly affect my agency, because my agency is purely about the kind of control I am able to achieve for myself regardless of what is happening to or around me.⁴⁰⁰ Further, recall here Nietzsche’s talk of the Sovereign Individual’s right to make promises. The right to make promises is afforded to the Sovereign Individual because he is able to master his own inclinations and thus, he is able to stand surety for his promises because of this motivational steadfastness. But there is also something else Nietzsche says about the Sovereign Individual, he claims in the *Genealogy* that:

To ordain the future in advance in this way, man must first have learned to distinguish necessary events from chance ones, to think causally, to see and anticipate distant eventualities as if they belonged to the present, to decide with certainty what is the goal and what is the means to it, and in general be able to calculate and compute. Man himself must first of all have become calculable, regular, necessary, even in his own image of himself, he is to be able to stand security for his own future, which is what one who promises does!⁴⁰¹

Nietzsche here claims that the Sovereign Individual has recognised the extent to which the external circumstances can undermine his ability to be certain that he will be able to reach the goal he has set for himself, or to fulfil the promise that he has made. Unless, Nietzsche *seems* here to suggest, he can remove those necessities and contingencies given by external circumstances, he is vulnerable to failure in his attempt to fulfil the promises he has made or to attain the goals he has set for himself. The Sovereign Individual looks, then, as if he might need to, like the Frankfurtian person, also only care about the kind of control he is able

⁴⁰⁰ This view might seem like an extreme, but what it is doing is providing us with an ideal picture – of course it is true that agency comes in degrees, and we may not be able to exercise this kind of freedom at all times. Torture and illness often break people, and the circumstances in which we find ourselves can certainly diminish our capacity (and strength of will) to exercise this kind of self-control. What Frankfurt, like Sartre, endorses is that this is the kind of freedom which is always available to us as human agents, though we may not always exercise it.

⁴⁰¹ GM II: 2.

to achieve for himself regardless of what is happening to or around him – that is, it looks as if the Sovereign Individual *might*, like the agent on Frankfurt's picture need to be self-sufficient: immune to the kinds of external influences which threaten to supplant her authority, at least over himself.

On the face of it then, both Frankfurt and Nietzsche's pictures are remarkably close to one other. On the Frankfurtian type picture, our capacity to exercise a kind of inner freedom through self-control is definitive of our agency. For Nietzsche, this is also true. And there is, of course, something quite significant about our capacity to exercise this kind of control over ourselves, to exercise the kind of inner freedom we take to be definitive of our agency. It is not surprising, then, that we spend a great deal of time reflecting on this capacity, thinking of ways to improve it, which will hopefully lead us to living lives which are more under our own control and less subject to the contingencies and necessities of the physical world in which we find ourselves – and I discussed some of these strategies which human agents might employ in chapter two. This line of reasoning, however, may further be suggestive of the idea that by gaining more control we will be able to live better lives – that is lives in which we exercise our agency well – *precisely because* our lives will be 'up to us', rather than determined by the forces which are external to us and which are indifferent to our well being. This line of reasoning has been suggested perhaps most fervently by the Stoics, and by various forms of asceticism, but I think it is also subtly suggested by Frankfurtian type pictures, as discussed above. While Nussbaum suggests that this is true of Nietzsche's picture of the Sovereign Individual, and we might be inclined to agree with her based on the above statements, in what follows I will argue that it would be a mistake to read Nietzsche in this way.

While I agree with Nussbaum that the self-emancipation characterisation of the Sovereign Individual provides us with, in many important ways, an ultimately unattractive ideal of strength at which to aim, I argue that this characterisation of the Sovereign Individual is itself problematic. It is my contention here that the Sovereign Individual, like the Stoic, is to be characterised in terms of his deep recognition of the necessity of his own vulnerability, but that, importantly, it is the Sovereign Individual's reaction to

this recognition that distinguishes the ideal of strength which we find in him from the problematic Stoic ideal. While the Stoics overemphasise the ideal subject's capacity for control, suggesting that he will do so in an attempt to transcend his necessary human vulnerability through an escaping and rejecting of it. It is, however, my contention that the Sovereign Individual is antithetical to the Stoic in precisely this respect: Sovereign Individuals live *through and with* their vulnerability by actively affirming it. The Sovereign Individual is thus, more properly, I argue, to be understood as embodying and affirming precisely the kind of fragility and vulnerability, which the Frankfurtian and Stoic person seeks to transcend through rejection and denial.⁴⁰² Given Nussbaum and Nietzsche's criticisms of the Stoic position, I argue that Nietzsche and Nussbaum may have more in common than Nussbaum suggests. And further, that in his depiction of the Sovereign Individual, Nietzsche, like Nussbaum, provides support for what I see as an epistemic normative prescription for personhood in the form of acknowledgement and affirmation.

Nussbaum has argued that the Frankfurtian line of reasoning outlined above is central in Ancient Greek philosophy to some extent, where Socrates, the Stoics and Aristotle all agree that we should "above all value our inner resources."⁴⁰³ Central to this line of reasoning is that, since the capacity for self-control definitive of agency is to be cultivated in order to make our lives go better, this is all that must be cultivated in order to truly live well. In the Stoic tradition the 'good person' is "a self-commanding person – one who, rather than being the slave of fortune, is truly free just because she doesn't care for the things that fortune controls. Commanding herself, she commands all that is important for living well; she is thus a person of real power and command in a world"⁴⁰⁴ where human vulnerability is to be overcome. Through exercising the capacity for self-

⁴⁰² In putting forward my own reading of the Nietzschean notion of the Sovereign Individual I do not necessarily take the notion to be either straightforward or uncontroversial. I recognise that there has been much debate amongst Nietzsche scholars about how we should interpret Nietzsche's notion. While I do not contrast my own reading of the notion of the Sovereign Individual with competing views in this literature explicitly in this paper, I do recognise that there may be room for contention and debate on this.

⁴⁰³ M. Nussbaum, 'Pity and Mercy', pp. 157-158.

⁴⁰⁴ *Ibid*, p. 146.

control, the Stoic gains power, and takes himself to have, thus, escaped his human vulnerability.

Nussbaum takes this Stoic line of reasoning, that having power over ourselves allows us power over “the vicissitudes of fortune,”⁴⁰⁵ to be central to Nietzsche’s account of the Sovereign Individual. She characterises the Sovereign Individual as one who is hostile to “human vulnerability and fragility in general,” seeing it as a kind of “impotence.” She quotes Nietzsche’s Aphorism 251 of *Daybreak* called ‘Stoical’ in which Nietzsche says:

There is a cheerfulness peculiar to the Stoic: he experiences it whenever he feels hemmed in by the formalities he himself has prescribed for his conduct; he then enjoys the sensation of himself as dominator.⁴⁰⁶

The suggestion is that by allowing ourselves to focus only on what is under our control we are able to remove all the chanciness and necessity that comes along with living in the physical world by making all the happenings and contingencies of that world no longer important for our well being. What happens inside of us is all that is relevant for here we have control in spite of what is or could be going on around us. Nussbaum argues that since Nietzsche’s approach is Stoic, his valorisation of self-command and self-overcoming can be criticised on precisely the same grounds that the Stoic’s can, namely, because of their failure to recognise that the vulnerability which is being escaped is in part necessary for living well. For Nussbaum, Nietzsche and the Stoics are “committed to denying that the physical goods of life are necessary conditions for *eudaimonia*. And thus... are committed to holding that people who are severely deprived, and even imprisoned and tortured, can still retain *eudaimonia*, so long as they are virtuous and self commanding.”⁴⁰⁷ She explains, I think quite convincingly, that the removal of the external conditions which make us vulnerable might be problematic because:

...one would need to decide how much worth persons and things and events outside ourselves actually have in the planning and conduct of our lives; what needs we actually have from the world and to what

⁴⁰⁵ *Ibid*, p. 151.

⁴⁰⁶ D 132 in *Ibid* in favour of this point.

⁴⁰⁷ *Ibid*, pp. 158-160.

extent those needs can be removed by a new attitude of self command toward and within oneself.⁴⁰⁸

However, she goes on to argue that the Nietzschean picture of ideal strength in the character of the Sovereign Individual is not an attractive picture of strength for precisely this reason, saying:

What should we think about the human being who insists on caring deeply for nothing that he himself does not control; who refuses to love others in ways that opens him to serious risks of pain and loss; who cultivates the hardness of self-command as a bulwark against all the reversals that life can bring? We could say, with Nietzsche, that this is a strong person. But there clearly is another way to see things. For there is a strength of a specifically human sort in the willingness to acknowledge some truths about one's situation: one's mortality, one's finitude, the limits and vulnerabilities of one's body, one's need for food and drink and shelter and friendship. There is a strength in the willingness to form attachments that can go wrong and cause deep pain, in the willingness to invest oneself in the world in a way that opens one's whole life up to the changes of the world, for good and for bad. There is, in short, a strength in the willingness to be porous rather than totally hard, in the willingness to be a mortal animal living in the world. The Stoic [and the Sovereign Individual]⁴⁰⁹ by contrast, looks like a fearful person, a person who is determined to seal himself off from risk, even at the cost of love and value.⁴¹⁰

In her criticism of Nietzsche, she goes on to say:

Nietzsche knows, or should know, this. For a central theme in his work is that Christianity has taught us bad habits of self-insulation and self-protection, alienating us from our love of the world and all of its chanciness, all of its becoming. On this account we have become small in virtue, and will remain small, unless we learn once again to value our own actions as ends, and our worldly existence as their natural home. I think that in the end Nietzsche fails to go far enough with this critique. He fails, that is, to see what the Stoicism he endorses has in common with the Christianity he criticizes, what "hardness" has in common with otherworldliness: both are forms of self-protection, both express a fear of this world and its contingencies...⁴¹¹

While I agree with Nussbaum that, were her characterisation of the Sovereign Individual correct, the Sovereign Individual would not provide us with an

⁴⁰⁸ *Ibid*, p. 156.

⁴⁰⁹ By extension because she takes him to be stoic, or even beyond stoic

⁴¹⁰ *Ibid*, p. 160.

⁴¹¹ *Ibid*.

attractive ideal at which to aim, because he would be living, as Nussbaum thinks, a radically impoverished human life by removing⁴¹² himself from the activities which, while on the one hand make us vulnerable, on the other actually add significant meaning and value to our lives. But Nussbaum does not go far enough with her claim that Nietzsche does, or ought to know, that this aspect of Stoicism is problematic.

For the Stoics, cultivating the capacity for self-control is an attempt to escape the contingency and vulnerability of a life lived in the physical world, which must be done through a kind of transcendence. This transcendence involves, for the Stoic, a rejection or a denial of the importance those aspects of our lives which are deeply vulnerable to the kinds of “contingencies and reversals,”⁴¹³ which Nussbaum suggests actively engaging with the world around us might bring. Isaiah Berlin has provided us with a canonical passage of what this line of reasoning amounts to, in which he describes a ‘retreat to the inner citadel’ in which we might take precisely this approach to transcendence in an attempt to gain control. He says:

I must *liberate* myself from desires that I know I cannot realize. I wish to be master of my kingdom, but my frontiers are long and vulnerable, therefore I contract them... to... *eliminate the vulnerable area*... The tyrant threatens me with imprisonment... But if I no longer feel attached to property, no longer care whether or not I am in prison... then he cannot bend me to his will... It is as if I had performed a strategic retreat into the inner citadel... I have withdrawn into myself; there and there alone, I am secure... *I illuminate obstacles in my path by abandoning the path*: I retreat to my own sect, my own planned economy, my own deliberately insulated territory, where... no external forces can have effect.⁴¹⁴

Berlin is hostile to this approach. In his discussion of freedom or autonomy, he rejects this line of reasoning about gaining control through the kind of transcendence suggested by the Stoics. As Berlin, I think rightly, argues, Stoic

⁴¹² The question here of removing oneself may be a question about values: by removing ourselves what we mean is that we no longer care about or value those things over which we have no control – for such a person these things are deemed of little value or worth.

⁴¹³ Here I am borrowing Martha Nussbaum’s terms, this issue is important for her in the context of living a flourishing life particularly. I am drawn to her conception of the flourishing life when assessing what the concept of a person is.

⁴¹⁴ I. Berlin, ‘Two Concepts of Liberty’, p. 129.

transcendence only appears to offer us freedom, only appears to offer us a path to follow in order to escape our vulnerability, but this path is deeply problematic. Berlin is particularly worried about how such a misconstrued picture of human freedom could be abused in political life, but Berlin's concerns, I think, highlight why even though Frankl's picture of the concentration camp inmate as still able to exercise agency, is plausible, it somehow riles against our intuitions that the inmate has been dehumanised – they have had an important aspect of their personhood undermined – and furthermore, as Berlin seems to suggest, why Frankl's picture riles against our intuitions that the camp inmate has been robbed of their ability to live well. He says:

If I find that I am able to do little or nothing of what I wish, I need only contract or extinguish my wishes, and I am made free. If the tyrant... manages to condition his subjects... into losing their original wishes and embracing... the form of life he has invented for them, he will, on this definition, have succeeded in liberating them. He will, no doubt, have made them *feel* free – as Epictetus feels freer than his master (and the proverbial good man is said to feel happy on the rack). But what he has created is the very antithesis...

If I save myself from an adversary by retreating indoors and locking every entrance and exit, I may remain freer than if I had been captured by him, but am I freer than if I had defeated or captured him? If I go too far, contract myself into too small a space, I shall suffocate and die. The logical culmination of the process of destroying everything through which I can possibly be wounded is suicide.⁴¹⁵

Nussbaum, similarly, rejects this kind of transcendence and valorisation of control, suggesting – as pointed out above – that we cannot live a good life by removing ourselves entirely from our entanglements and engagements with the world. Given my discussion of the role I see vulnerability playing in living well, I agree with Berlin and Nussbaum in their criticism of this Stoic transcendence.

Increasing our self-control in the way Frankfurt and Nietzsche suggest in their pictures, is important and vital for our understanding of ourselves as human agents. When control is central, however, we lose sight of an aspect of our personhood which I believe to be of the utmost importance to us – the necessity and centrality of human vulnerability which Nussbaum points out most explicitly in her work (not only the paper discussed here, but throughout *The Fragility of*

⁴¹⁵ *Ibid.*, p. 164.

Goodness as well). There is an aspect of our personhood which seems to contrast most explicitly with the capacity for agency understood in terms of self-control, that is, our reliance as persons on external goods and those things over which we do not have complete control. While it seems clear that agency is necessary to our understanding of what it means to be a person, to be a person also requires a recognition on the part of the person themselves of their own limits and vulnerability – the ability to recognise the extent to which circumstances, events and other persons play a role “in the planning and conduct of our lives”⁴¹⁶ and contribute to our ability to live well. If our understanding of what is most central to our conceptualisation of what we as human persons are is somehow at odds with what is central for achieving or maintaining good or worthwhile lives, then I think we have done a great injustice to the notion of personhood, and we face a great danger because of this. We face a great danger because a misconstrued picture of what is central to personhood will lead us to developing an ideal, a picture of personhood towards which we ought to strive, and if our ideal is out of sync with what we actually take to be important for living well, then we will have a great deal of difficulty achieving a good life by aiming at such an ideal. *Contra* Nussbaum, I think that Nietzsche recognises this, and addresses this issue when discussing his ideal of personhood made manifest in the Sovereign Individual. I see Nietzsche, similarly to Nussbaum, providing support for a normative prescription for personhood in the form of acknowledgement and affirmation of our vulnerability.

Why might Nussbaum think of the Sovereign Individual as offering us with a Stoic ideal at which to aim? Well, Nietzsche does say that:

Honesty, supposing that this is our virtue from which we cannot get away, we free spirits – well, let us work on it with all our malice and love and not weary of “perfecting” ourselves in our virtue, the only one left us... And if our honesty should nevertheless grow weary one day and sigh and stretch its limbs and find us too hard, and would like to have things better, easier, tenderer, like an agreeable vice – let us remain hard, we last Stoics!⁴¹⁷

⁴¹⁶ M. Nussbaum, ‘Pity and Mercy’, p. 156.

⁴¹⁷ BG 227.

And here it would be easy to misinterpret what Nietzsche says because of his explicit reference here to Stoicism. In fact, Nietzsche often says things which we may interpret as supporting the Stoic doctrine of asceticism, especially when he says like:

To those human beings who are of any concern to me I wish suffering, desolation, sickness, ill-treatment, indignities – I wish that they should not remain unfamiliar with profound self-contempt, the torture of self-mistrust, the wretchedness of the vanquished: I have no pity for them, because I wish them the only thing that can prove today whether one is worth anything or not – that one endures.⁴¹⁸

And citing this passage, Nussbaum argues that Nietzsche “does not grasp the simple fact that if our abilities are physical abilities they have physical necessary conditions,” he does not grasp what she calls a “basic vulnerability,” and that this leads Nietzsche to his conclusion “that even a beggar can be a Stoic hero.”⁴¹⁹ And so Nussbaum interprets this passage of Nietzsche as aligning him with the Stoic ideal of transcendence. Certainly, there are some aspects of Nietzsche’s account of morality and personhood which are vulnerable to the criticisms that Nussbaum raises here. However, there is another aspect of Nietzsche’s account of personhood that is not, namely his suggestion that the Sovereign Individual not only acknowledges, but affirms his vulnerability. It is this aspect of Nietzsche’s account which I think requires closer attention, and an aspect of Nietzsche’s account which I think Nussbaum should be sympathetic to.

When Nietzsche talks about the Sovereign Individual’s right to make promises he emphasises that the right to make promises is an act of self-overcoming. For Nietzsche, however, this overcoming cannot be seen, as the Stoic would have us believe, as a ‘retreat to the inner citadel’⁴²⁰ by which we deny the important role those aspects of our lives that are not under control play in our living well. In order to have the right to make promises, we must also recognise the important role our own vulnerability plays in the actual planning and conduct of our lives; as we have seen this is necessary for the Sovereign

⁴¹⁸ WP 910.

⁴¹⁹ M. Nussbaum, ‘Pity and Mercy’, pp. 158-160.

⁴²⁰ See I. Berlin, ‘Two Concepts of Liberty’.

Individual to have the right to make promises at all. That we are deeply vulnerable is not just something that the Sovereign Individual can ignore. For Nietzsche unless we can be honest with ourselves about the role of things which are beyond our control (that is if we attempt to transcend our vulnerability) we would be guilty of *ressentiment*.⁴²¹ It is precisely this honesty with ourselves about our own 'all too human' condition that Nietzsche thinks we should cultivate and remain steadfast in, in order to prevent ourselves from succumbing to *ressentiment*. What is beyond our control is vital precisely because it forms an important part of who we actually are – unlike the man of *ressentiment*, in owning up to this realisation the Sovereign Individual has a more integrated understanding of herself which includes not only those aspects of herself which 'command' (are under my control) but also those which 'obey' (which are beyond my control). What this shows us is that for Nietzsche, as Elveton puts it, "[t]he fundamental Stoic opposition between what is mine (my will and what falls under its direct control) and not mine reduces the self in a one-dimensional and artificial way."⁴²²

But perhaps even more importantly for my argument against Nussbaum's characterisation of the Sovereign Individual, Elveton claims that Nietzsche rejects the fundamental Stoic picture in which it "is my attitude, my inner composure, that is reflective of my individual power.... [and so] my actions in the world elude me and are not a significant part of me... what I am is not so much what I do, but my rational attitude toward what I do, and my rational attitude toward what is done to and what happens to me."⁴²³ I agree with Elveton that Nietzsche is against this, and I think that this explains Nietzsche's claim that we cannot separate the doer from the deed for precisely this reason. We might like to think here of the seemingly strange account of personhood in the 13th section of the *Genealogy* where Nietzsche admonishes the separation between the 'doer' and the 'deed'. This passage suggests that Nietzsche is reluctant to view personhood as something which could be separated from our actual actions in

⁴²¹ As pointed out previously, for Nietzsche, *ressentiment* is, at least in part, a psychological condition which results in a fractured and damaged individual, lacking in integrity.

⁴²² R. Elveton, 'Nietzsche's Stoicism: The Depths are Inside', in Bishop, P., (ed.), *Nietzsche and Antiquity*, Rochester, Camden House, 2004, p. 195.

⁴²³ *Ibid.*

the world of riskiness and chance. Nietzsche claims that there is an inextricable link between persons and their actual experiences, saying also: “if I remove all the relationships, all the properties, ‘all the activities’ of a thing, the thing does not remain over”. Moreover, I think, that Nietzsche cannot be seen to valorise self-control in the Stoic mode of transcendence precisely because, above all, Nietzsche wants us to affirm life, ourselves *and* the world of chance and necessity in which we live. This is the world we live in and we cannot seek to escape it, but must rather seek to thrive in it, and this will require our recognition of our vulnerable place in it. Above all in his Doctrine of Eternal Recurrence, where according to Nietzsche, “nothing that has happened to us is contingent,”⁴²⁴ and affirming any given aspect of our lives or selves entails our affirming all aspects of our selves, our pasts, and indeed the whole history of the physical world in its entirety. Recall Nietzsche’s description of the Doctrine of Eternal Recurrence in *The Gay Science*:

‘This life as you now live it and have lived it you will have to live once again and innumerable times again; and there will be nothing new in it, but every pain and every joy and every thought and sigh and everything unspeakably small or great in your life must return to you, all in the same succession and sequence... The eternal hourglass of existence is turned over again and again, and you with it speck of dust!’⁴²⁵

Through the Doctrine of Eternal Recurrence Nietzsche claims most explicitly that the harmonious integration of various aspects of ourselves required for self-creation entails our affirmation of all of these aspects. According to Nehamas:

The final mark of this integration, its limiting case, is provided by the test involved in the thought of the eternal recurrence. This mark is the desire to do exactly what one has already done in this life if one were to live again.⁴²⁶

Nietzsche suggests in *The Gay Science* that there can only be two possible reactions to this psychological test posed by the doctrine: that we reject it as the

⁴²⁴ See A. Nehamas, ‘The Eternal Recurrence’, p. 123.

⁴²⁵ GS 341.

⁴²⁶ A. Nehamas, ‘How One Becomes What One Is’, p. 274.

most detestable malison or we welcome it with the greatest joy. Nietzsche writes:

Would you not throw yourself down and gnash your teeth and curse the demon who spoke thus? Or have you once experienced a tremendous moment when you would have answered him: 'You are a god, and never have I heard anything more divine.'⁴²⁷

What is important for our reaction to the suggestion of the doctrine is the understanding that it will be all aspects of our life, our self, and indeed the entire chronological history of the world as it has been, that we will be either affirming or bemoaning. Recall that, for Nietzsche, if we deny even the smallest part of who we are or what has come before, we cannot affirm our present selves, for our present self is necessarily constituted by our own past and all effects of the world on it. What Nietzsche wants us to realise is that if we can at any point in our lives affirm who we are, even for a 'moment', we must necessarily affirm all aspects of ourselves, our past actions, attitudes and opinions. "Have you ever said Yes to a single joy? O my friends, then you have said Yes to all woe,"⁴²⁸ proclaims Zarathustra. And considering this carefully, we understand that in order to answer positively in light of the demon's question, to affirm the demon who presents us with the doctrine, we must will all that has gone before, even the very worst of the worst: and in realising this, Nietzsche's ultimate man, Zarathustra, "finally becomes able to want to undergo again all that is cheap and detestable about the world for the sake of what is not."⁴²⁹ Through the doctrine of eternal recurrence, Nietzsche "asks us whether we merely want to drift with the tide of things or whether we would be creators"⁴³⁰ – whether we would float along unthinkingly or whether we would engage our capacity to actively affirm all that is past, all that is present, and indeed all that is necessary. "Prior to [this, Nietzsche asks in the thought of the eternal return] whether we desire *the conditions* by which we might again become creators."⁴³¹

⁴²⁷ GS 341.

⁴²⁸ Z IV 19.

⁴²⁹ See A. Nehamas, 'The Eternal Recurrence', p. 124.

⁴³⁰ M. Heidegger, *Nietzsche; Volume 2: The Eternal Recurrence of the Same*, Krell, D., trans., New York, Harper & Row Publishers, 1984, p. 174.

⁴³¹ *Ibid*, p. 174.

Nietzsche suggests that the Sovereign Individual expresses an inner harmony grounded in the affirmation of life - such as that which is involved in affirming the demon's proposition of the eternal recurrence of the same. This affirmation and revaluation is required for the kind of self-creation which is compatible with Nietzsche's *Amor Fati*. As I have already argued, a plausible explanation of what he means here is that we affirm, along with all the joy and all the positive aspects of ourselves, the suffering and contrary aspects of ourselves as well. The acceptance, or rather love, of fate adopted by the Sovereign Individual, then, is also to be understood as a normative requirement for personhood. We must affirm our lives in their particular structure and in their entirety. In section 49 of the *Twilight of the Idols*, Nietzsche says:

Such a spirit who has become free stands amid the cosmos... joyous and trusting... in the faith... that all is redeemed and affirmed in the whole.⁴³²

In his account of personhood Nietzsche seems to go further than simply saying that we are required to acknowledge the conflicting aspects of ourselves which constitute our whole-selves - as discussed in the previous chapter - he claims that we need to actively affirm these aspects. A word of clarification on the sense in which Nietzsche here uses the term affirmation is required. As I have already suggested, there are parts of ourselves which we do wish to overcome, and this overcoming requires knowledge of what is to be overcome as well as the knowledge that what is to be overcome is precisely some aspect of my very own self. Here Nietzsche's use of the term affirmation might sound misleading - why should I affirm that aspect of myself which I wish to overcome? What I take Nietzsche to mean by this is that I must affirm those aspects because, on the one hand, they are my own and on the other hand, because without those aspects I would have nothing to overcome and my ability to overcome them is the basis for my ability to exercise self-governance in the first place. Importantly, for Nietzsche, there will be times when certain aspects of ourselves, the more deeply ingrained aspects of ourselves, cannot be overcome; there are, as Frankfurt would say, recalcitrant aspects of ourselves. For Nietzsche, this is also a

⁴³² TI 49.

fundamental part of what it means to be a person – the ‘animal’ man. In coming to terms with this Nietzsche suggests that it is of primary importance not only to affirm these aspects of ourselves – reinterpreting their value for ourselves – but also to affirm in this respect the kinds of creatures we humans are.

Nietzsche’s doctrine suggests that if we deny even the smallest part of who we are or what has actually happened in the world, we cannot affirm our present selves, for our present self is necessarily constituted by our own past and all effects of the world on it – we cannot separate who we are from our lives, nor can we separate our lives from the world in which they have been lived. So, for Nietzsche, although “Stoic thought is suggestive of spiritual strength” to some extent, he also sees it as “superficial, with fateful consequences.”⁴³³ He says most tellingly:

I believe that we do not understand Stoicism for what it really is. Its essential feature as an attitude of the soul... [a]... comportment toward pain and representations of the unpleasant: [it is] an intensification of a certain *heaviness* and *weariness* to the utmost degree in order to weaken the experience of pain. Its basic motifs are *paralysis* and *coldness*; hence a form of anaesthesia.... *In summa: turning oneself into stone* as a weapon against suffering and in the future conferring all worthy names of divine-like virtues upon a statue... I am very antipathetic to this line of thought. It undervalues the value of pain (it is as useful and necessary as pleasure), the value of stimulation and suffering. It is finally compelled to say: everything that happens is acceptable to me; nothing is to be different. *There are no needs over which it triumphs* because it has killed the passion for needs.⁴³⁴

Broadly speaking, Stoicism is for Nietzsche a doctrine in which we acquire self-salvation by transcending or escaping the world in which we live, and this is precisely the kind of anti-naturalism which Nietzsche is at pains throughout his works to rally us against. Although the Stoics purport to be naturalists, Nietzsche thinks that they offer a new brand of anti-naturalism by falsely transposing their ideals on nature. Nietzsche says in *Beyond Good and Evil*:

“According to Nature you” want to *live*? O you noble Stoics, what deceptive words these are! Imagine a being like nature, wasteful beyond measure, indifferent beyond measure, without purposes and

⁴³³ R. Elveton, ‘Nietzsche’s Stoicism’, p. 193.

⁴³⁴ F. Nietzsche in *Ibid*, p. 200.

consideration... -how *could* you live according to this indifference?... In truth, the matter is altogether different: while you pretend rapturously to read the canon of your law in nature, you want something opposite... Your pride wants to impose your morality, your ideal, on nature.⁴³⁵

What a more accurate version of naturalism, according to Nietzsche, teaches us is precisely that vulnerability, contingency and risk are part of this world and so also our lives in it. In seeking to transcend this aspect of the world and our lives we strive towards what he refers to as 'other-worldly' hopes, and cannot properly affirm this life, this world, as it is – and this will impoverish our experience of it. For Nietzsche, Elveton argues:

Stoic morality testifies to a very high level reached by man's moral consciousness, but in the last resort it is a decadent and pessimistic morality of despair, which sees no meaning in life; it is inspired by the fear of suffering. One must lose sensitiveness to suffering and become indifferent – that is the only way out.⁴³⁶

For Nietzsche, this is unacceptable, and for this reason he characterises his hero, the Sovereign Individual, against this Stoic ideal as having “the strength to suffer pain and to *add to it*.”⁴³⁷ Unlike the Stoic, then, the Sovereign Individual does not seek to escape the misery and pain which comes along with being vulnerable to all the chance and necessity which the world holds in store for us, but seeks to actively affirm it, live through and with it. The Stoics cannot affirm life to the highest degree precisely because they seek to transcend their vulnerability and the suffering of this world, and for Nietzsche, this is symptomatic of the 'will to nothingness' – a form of nihilism seen in “Platonism-late Judaism, Christianity and 'slave morality'.”⁴³⁸ As May elegantly explains:

They will 'nothing' because they are driven by an all-consuming will to escape a world of suffering, a will that, because it repudiates what is constitutive of living – the loss or elusiveness of what we most desire, such as loved ones, health, achievements, predictability, joy, and ultimately life itself – wills what is not human life, not the world of transience, chance, fate, and time in which we are actually situated. In refusing to affirm that life is

⁴³⁵ BG 9.

⁴³⁶ R. Elveton, 'Nietzsche's Stoicism', p. 199.

⁴³⁷ *Ibid*, p. 200.

⁴³⁸ See S. May, 'Nihilism and the Free Self', in Gemes, K., and May, S., (eds.), *Nietzsche on Freedom and Autonomy*, Oxford, Oxford University Press, 2009, p. 89.

structured by the possibility of loss they imagine an ideal order out of which this possibility has been conceptually airbrushed, an order that is clearly not the one into which humans are born.⁴³⁹

Thus, it is only by affirming his own vulnerability that the Sovereign Individual is able to engage in 'affirming life to the highest degree' and thus living well.

Wallace has summarised Nietzsche's position in the Doctrine of Eternal Recurrence as follows:

[T]he idea of eternal recurrence is tied to our epistemic situation. This situation is one in which we have excellent reason to believe that our unconditional affirmation of our own lives commits us to affirming all manner of regrettable events and circumstances in the past, but without our knowing of any particular event or circumstance that we are implicated in it in particular [...] only if we are prepared to will the totality of world history can we honestly adopt an attitude of unconditional affirmation toward our lives and the things to which we are attached. Because for all we know, that attitude already commits us to affirming as well the most catastrophic and egregious aspects of the larger histories in which our lives are caught up.⁴⁴⁰

One major objection to the Nietzschean suggestion of affirmation, discussed by Wallace, is that unconditional affirmation of our lives entails affirmation of events and states of affairs which do not seem to warrant our affirmation. If we affirm all of history in affirming ourselves, we affirm catastrophes, wars, genocides, slavery, inequality, and innumerable other states of affairs for which affirmation would seem, at best, inappropriate. The unconditional Nietzschean affirmation, according to this criticism, requires that we affirm things which we ought not to affirm if we unconditionally affirm our lives. Wallace concludes that in maintaining an attitude of unconditional affirmation we affirm that it is all things considered better that we have lived our lives, rather than not having lived at all, even though this requires us to also affirm the existence of things which considered in isolation do not warrant our affirmation:

But what about this attitude of unconditional affirmation itself? Couldn't we revise this attitude, adopting toward our lives a more ambivalent stance once we are clear about our potential implication in historical catastrophes

⁴³⁹ *Ibid*, p. 89.

⁴⁴⁰ R. Wallace, *The View from Here*, p. 253.

of various kinds? This is a theoretical possibility – there are people, after all, who give up on life while they are in the midst of it, and they would not appear to affirm their lives unconditionally. But this is not how things are for most of us. The same vital forces that lead us to cling to life as we are living also give rise to an unconditional preference to have lived – a preference, looking backward, to have lived our actual lives, as against the alternative that we should not have lived at all [...] Our plight as humans might be that we are condemned, in virtue of our attachment to life, to affirming conditions in the world that we cannot possibly regard as worthy of this attitude. There is something absurd about this situation, involving as it does the persistence of attitudes that don't fundamentally make sense to their bearers. But the absurdity may be one that is endemic to the human condition.⁴⁴¹

And given my reading of Nietzsche above, Wallace's sensible conclusion here is something that Nietzsche himself is likely to accept, given that a central aspect of Nietzsche's account which I highlighted above is that we ought, as persons, to recognise and affirm our own vulnerability, and in so doing, unconditionally affirm the life that we have lived. Part of affirming our vulnerability in this way must, I think, involve our recognition of what Wallace calls an absurd situation, that human vulnerability and the undeniable value of a human life lived well go hand in hand with one another. To wish for a life devoid of vulnerability, devoid of things which we may have difficulty affirming, is to wish for a life that is no longer the life lived by a person like ourselves.

My life, I think, would not be recognisably human *and* it would be radically impoverished, if not wholly undesirable, if I did not care very deeply about the things over which I have no control – specifically we could mention concerns such as how my projects actually fare in the external world and my interpersonal relationships with other people. Here, I think, that Nussbaum is right to point out that our human vulnerability and our reaction to this vulnerability are necessary for understanding what good agency, or personhood, entails. In this section, I have argued following Nussbaum and Nietzsche, against the Frankfurtian and Stoic normative pictures of personhood, claiming that it is an epistemic normative requirement for personhood that we recognise and affirm our own vulnerability.

Taking the notion of vulnerability as a fundamental idea in Nietzsche's ethics may, at first blush, seem somewhat controversial, and it has certainly not

⁴⁴¹ *Ibid.*

been a widely discussed aspect of Nietzsche's philosophy in general. As I have argued in this section, I take the notion of vulnerability to be something addressed by Nietzsche in subtle, yet nuanced ways. I think that this aspect of Nietzsche's philosophy warrants further philosophical investigation. In the following section I argue for a final normative prescription for personhood in the form of authenticity.

4.3. Authenticity and the Art of Living

In the hit song 'My Way', written by Paul Anka and most popularly performed by Frank Sinatra, questions about human agency, personhood and authenticity are simultaneously raised in the following lyrics:

*I planned each chartered course,
each careful step along the byway,
and more, much more, than this,
I did it my way...⁴⁴²*

The first two lines in the quoted lyrics tap into a picture of human agency which I explored in detail in the first part of this thesis. Human agents, on such a view, are able to construct certain life plans and goals within the framework of liberty and freedom accorded by their ability to exercise self-control. Rachels and Ruddick have argued that our agency, understood in this way, is a necessary constitutive part of personhood:

freely chosen actions are constitutive elements of one's life, while forced actions are interruptions, limitations, or obstacles in that life... By exercising these capacities in making choices... [we] create... [our] lives: liberty makes lives not beautiful, but possible.⁴⁴³

I agree with Rachels and Ruddick here, that agency is a fundamentally constitutive part of our even being able to set goals, make life plans, and, ultimately, to live the kind of life we consider human.

⁴⁴² My Way, P. Anka, writer, F. Sinatra, performer, Reprise, Los Angeles, 1968.

⁴⁴³ W. Ruddick, and J. Rachels, 'Lives and Liberty', in Christman, J., (ed.), *The Inner Citadel: Essays on Individual Autonomy*, Oxford, Oxford University Press, 1989; p. 230.

But what is meant by Sinatra when he sings that “more than this”, he did it his way? How can we explain the notion of, over and above being agents and thus being in control of our lives in the relevant sense, also living our lives in a manner which is uniquely our own, or *our way*? It is in answering these questions, I think, that we are prompted to consider what it means to live well in a way not often discussed in analytic philosophical accounts of personhood such as Frankfurt’s. While human agency makes our lives possible, I will argue in what follows that when we live authentically, our lives are made not only possible, but also beautiful. To live well, I will argue, is to live authentically, and so authenticity can be seen as a normative requirement for personhood.

Documents or objects are considered to be authentic when we can verify their origins. Authentic documents and objects are originals in the sense that they are of undisputed origin or are considered genuine. When we refer to authentic persons, however, the question of origins is not typically what is at stake.⁴⁴⁴ When referring to persons we typically use the notion of authenticity to refer to what can be seen as an *ideal mode of existence*, one in which a person lives their life in such a way that they might be considered to have been true to themselves. In *The Gay Science*, Nietzsche suggests that in order to be true to ourselves, in order to become Sovereign Individuals, we must listen to our conscience which shouts, “You must become who you are.”⁴⁴⁵ This idea of being true to oneself, or being authentic, which lies at the very heart of Nietzsche’s understanding of the Sovereign Individual, requires further explanation, since the idea of being true to oneself, or authentic, has been understood in a number of ways.

According to Charles Guignon in his popular book *On Being Authentic*, the “contemporary ideal of authenticity directs you to realize and be that which you already are, with the unique, definitive traits already there within you.”⁴⁴⁶ According to this ideal of authenticity, we already have a defining essence buried deep within us, and it is only in being true to this essence – by giving expression

⁴⁴⁴ We might well imagine science-fiction scenarios of human cloning or thought experiments of brain-body swaps, in which determining the origin of a person would be important for determining their authenticity, however, we refer to persons as authentic in everyday life in an altogether different sense.

⁴⁴⁵ GS 270.

⁴⁴⁶ C. Guignon, *On Being Authentic*, London and New York, Routledge, 2004, p. 3.

to all the traits and dispositions which are already there within us – that we can exercise our authenticity. From literature to self-help books the words of Shakespeare’s Polonius ‘To thine own self be true’⁴⁴⁷ have been invoked in support of this view of authenticity. Cooper thus refers to this view as the ‘Polonian View’.⁴⁴⁸ He explains that on such a view what is required for authenticity is that we turn our gaze inwards on ourselves and discover who we really are. Who we most truly are, is thus assumed to already lie within us, identifiable through self-reflection and in a sense ready-made. The Polonian view thus understands authenticity as a project of my uncovering and expressing what is already there within me. Similarly, on Frankfurt’s account, when the authoritative power of our better judgement shines forth in our behaviour and actions, we exercise the self-governance associated with personhood. Frankfurt, however, claims that rather than simply uncovering a pre-existing true self lying deep within ourselves, we must play an active role in the formation or constitution of ourselves – we must be involved in some form of self-control in which we are responsible not only for self-modification, but self-creation. On his account a request for active control in the form of self-creation, rather than merely self-discovery, is central to securing our understanding of ourselves as particular persons. Nietzsche finds similar fault with the Polonian view of authenticity, according to Cooper:

Nietzsche is, first, consistently hostile to the metaphysics of the self encouraged by Polonian talk: to the idea of a self which persists, unchanged in its essence, through the vicissitudes of life, and hidden perhaps from view. Such a self is as mythical for Nietzsche as it had been for Hume; hence his references to the ‘false substantializing of the “I”’, and to the wrong-headed view that the self is the source of thought when, on the contrary, it is ‘through thought that the “I” is posited’.⁴⁴⁹

Nietzsche agrees that a form of self-creation is required for the project of becoming who one is. In *The Gay Science* Nietzsche claims:

⁴⁴⁷ W. Shakespeare, *Hamlet*, United Kingdom, Penguin Classics, 1994, Act 1, scene 3, 78.

⁴⁴⁸ D. Cooper, *Authenticity and Learning: Nietzsche’s Educational Philosophy*, Great Britain, Routledge and Kegan Paul plc., 1991, p. 8.

⁴⁴⁹ *Ibid*, pp. 13-14.

We, however, want to become those we are – human beings who are new, unique, incomparable, who give themselves laws, who create themselves.⁴⁵⁰

Much like Frankfurt's persons, Sovereign Individuals, on Nietzsche's account, are to be thought of as actively asserting control over or governing themselves in such a way as to make their actions and lives their own, and which establish them as particular someone's. According to Pippin on Nietzsche's account: "If herd morality, conformism and sheep-like timidity are to be held in contempt, then some contrary notion seems suggested, some ideal of social independence and a kind of self-rule or self-reliance."⁴⁵¹ Emerson, in his famous piece "Self-Reliance," makes the following rather dramatic claim:

Society everywhere is in conspiracy against the manhood of every one of its members. Society is a joint-stock company, in which the members agree, for the better securing of his bread to each shareholder, to surrender the liberty and culture of the eater. The virtue in most request is conformity. Self-reliance is its aversion [...]. Absolve yourself to yourself, and you shall have the suffrage of the world.⁴⁵²

Like Emerson, Nietzsche's idea of personhood is intricately linked to the idea of challenging blind conformity to the values and ideals of society at large. Nietzsche suggests that by blindly conforming to society's values and ideals, we deny our capacity to derive our values for ourselves, which is fundamental to our living well. As much is evident when Nietzsche claims the following in *Schopenhauer as Educator*:

The man who would not belong to the mass needs only to cease being comfortable with himself; he should follow his conscience which shouts at him: 'Be yourself; you are not really all that which you do, think, and desire now.'⁴⁵³

⁴⁵⁰ GS 338.

⁴⁵¹ R. Pippin, 'How to Overcome Oneself', p. 76.

⁴⁵² R. Emerson, 'Self-Reliance', in *Essays and Lectures*, New York, Library of America, 1983, p. 261.[Sic.]

⁴⁵³ SE in Hollingdale, R., trans., *Friedrich Nietzsche: Untimely Meditations*, Cambridge, Cambridge University Press, 1989, p. 127.

For Nietzsche, subscribing to, or endorsing, the values and ideals advocated by society, or any other source of authority for that matter, poses a danger to the possibility of our becoming Sovereign Individuals. The biggest danger, however, is that we may find ourselves having slipped into an unreflective acceptance of these values and ideals.⁴⁵⁴ Nietzsche's claim that we should be self-creating agents or Sovereign Individuals can thus, first and foremost, be seen as a call to reflect not only on our social existence, but the values and ideals which lie at the core of this existence. Moreover, according to him, by unreflectively accepting transmitted values and ideals, we might be led to make equally unreflective assessments and evaluations about aspects of ourselves. Thus, we may find ourselves slipping into comfortable unreflective understandings of our selves. For Nietzsche, this is a most dangerous threat to our sovereignty and our ability to truly exercise our agency well. In *Schopenhauer as Educator* Nietzsche claims:

The man who would not belong to the mass needs only to cease being comfortable with himself; he should follow his conscience which shouts at him: 'Be yourself (*sei du selbst*); you are not really all that which you do, think, and desire now'.⁴⁵⁵

What is most important for our authenticity, however, is a kind of narrative integrity. Recall here my discussion of narrative integrity in the previous chapter. Recall that for Nietzsche our lives and indeed our selves bear striking resemblance to works of fiction or biography in that they rely on a similar kind of narrative structure. Since our lives resemble, for Nietzsche, the lives of literary characters in their structure and composition, the kind of inner harmony or internal coherence with which we associate the authentic individual is a kind of narrative integrity. Recall further, Nehamas suggests that Nietzsche's concern with authenticity is rooted in the fact that much of Nietzsche's thinking revolves around literary models. However, Nietzsche's thinking, almost as often as centring on literary models, centres on artists and the creative endeavour of art. Nietzsche views art as not merely a form of human activity, but as the highest expression of what the human spirit is capable. Most notably in *The Birth of Tragedy*, Nietzsche claims

⁴⁵⁴ See Cooper, *Authenticity and Learning*, p. 4.

⁴⁵⁵ SE in Hollingdale, R., *Friedrich Nietzsche*, p. 127.

that 'art, and *not* morality, is presented as the truly metaphysical activity of man.'⁴⁵⁶

The importance of authenticity to our personhood on Nietzsche's account can be understood when we think of what is wrong with an artistic forgery, and consequently why it is that we value those masterful artists who have shaped the history of Art itself. According to Alfred Lessing, we can understand the notion of a forgery in light of its antithetical occurrence, which encapsulate what we mean by 'genuineness' or 'authenticity':⁴⁵⁷ "forgery is a concept that can be made meaningful only by reference to the concept of originality, and hence only to art viewed as a creative, not as a reproductive or technical activity."⁴⁵⁸ He claims, moreover, that the offensiveness of forgery lies not in the fact that it is "against the spirit of beauty (aesthetics) or the spirit of law (morality),"⁴⁵⁹ but rather that is "against the spirit of art"⁴⁶⁰ itself. What this means then when we are considering the art of living (as the creative response of the individual to the world), is that without the kind of originality which is lacking in the artistic forgery our lives are not merely against the 'spirit of art', but against the very spirit of life itself. To have a life devoid of opportunity to respond creatively to our environment encompassed by the normative requirement for authenticity suggested here, is to have a life in which all activity lacks artistic integrity in the same way that the artistic forgery lacks artistic integrity, and prevents our lives from being beautiful. Authenticity, for Nietzsche, is thus to be seen as a kind of artistic integrity about our lives.

Nietzsche frequently implores us to take the artist's perspective on our lives. Nietzsche thinks that in shaping our lives in the way that great artists compose their works we can truly begin to understand the convergence of necessity and creation – that is, express *Amor Fati*. Nietzsche claims that:

We should learn from artists while being wiser than they are in other matters. For with them this subtle power [of arranging things and of making them beautiful] usually comes to an end where art ends and life

⁴⁵⁶ BT 5.

⁴⁵⁷ A. Lessing, 'What is Wrong with a Forgery?', in Warburton, N., (ed.), *Philosophy: Basic Readings*, 2nd Edition, London and New York, Taylor and Francis, 2005, p. 536.

⁴⁵⁸ *Ibid*, p. 543.

⁴⁵⁹ *Ibid*, p. 542.

⁴⁶⁰ *Ibid*.

begins; but we want to be the poets of our life – first of all in the smallest, most everyday matters.⁴⁶¹

Nietzsche claims that the unity of self required for living authentically is found primarily in artists, who he claims:

seem to have more sensitive noses in these matters, knowing only too well that precisely when they no longer do something ‘voluntarily’ but do everything of necessity, their feeling of freedom, subtlety, full power, of creative placing, disposing and forming reaches its peak – in short, that necessity and ‘freedom of will’ then become one in them.⁴⁶²

Nietzsche praises those who have become who they are by fashioning their lives as the great artists have done, with their own individual style. Artists who are seen by Nietzsche to own themselves through the expression of their individual style, experience on Nietzsche’s view what we might be inclined to call ‘the art of living’. Thus, I refer to the final feature of the Sovereign Individual on Nietzsche’s picture as ‘individual style’. The introduction of the notion of individual style introduces into our evaluations of persons a consideration of whether a person’s actions and behaviours make up a personality or an identity which could be judged aesthetically. This consideration, it seems to me, is relevant to our understanding of ourselves as persons. In fact, this is one of the primary considerations which inform our interactions with one another as persons with individual personalities capable of authentic interactions and lives.⁴⁶³

What traditional accounts of agency and personhood try provide us with are the conditions under which we can make our lives our own, but what they do not explain is that, over and above simply making our lives our own, we also want to make our lives beautiful: we want to be engaged in the fine art of living. Nietzsche’s Sovereign Individual is not simply free from constraints, but is free to shape her life with the aesthetic qualities which in a very important sense make that life worth living. The account of Nietzsche’s Sovereign Individual, as I have explained it in this chapter, it is concerned with our directing ourselves to “use

⁴⁶¹ GS 299.

⁴⁶² BG 213.

⁴⁶³ See A. Nehamas, ‘How One Becomes What One Is’, p. 276.

and interpret experience in [our] own way"⁴⁶⁴ – to become who we are artistically. This view of authenticity thus adds aesthetic value to our lives which traditional accounts of personhood, such as Frankfurt's, fail to capture.

⁴⁶⁴ J. Mill, *On Liberty*, p. 122.

Conclusion

I think... that I would rather recollect a life mis-spent on fragile things than spent avoiding moral debt... I wondered just what I had meant by fragile things.

There are so many fragile things, after all. People break so easily, and so do dreams and hearts.⁴⁶⁵

In this thesis I have argued for distinct conceptions of human agency and personhood which incorporate the important consideration of our own human vulnerability. While I defended the self-control view of human agency central to analytic philosophical accounts such as Frankfurt, Watson and Mele's, I have supplemented this account arguing that vulnerability is also constitutive of human agency and personhood. I have also argued that personhood should be understood as a substantive notion pointing to our ability to exercise our agency well, or to thrive. Following Neil Gaiman, quoted in the epigraph above, I see human agency and personhood as 'fragile things'. Like Nussbaum, however, I also see this vulnerability as an essential ingredient for our thriving.

Against Frankfurt's normative prescription for personhood in the form of wholeheartedness, I have offered four normative prescriptions for personhood of my own. While there may be a number of further normative prescriptions for personhood, I see the ones presented here as offering some hope for the possibility of living a meaningful and worthwhile human life in the face of our vulnerability.

As persons, I have argued, we must live our lives embedded in the world; embedded in meaningful interactions with others; and living our lives to the fullest will require pursuing goals and ends which may result in failure and loss. In order to exercise our agency well, then, we are required to make the sincere acknowledgment of our own vulnerability. That is, as Pedro Tabensky points out, "the sincere acknowledgement that life is largely about muddling through darkness, always with limited personal resources at one's disposal."⁴⁶⁶ This vulnerability, I have argued, is neither something to always be escaped, nor always something

⁴⁶⁵ N. Gaiman, *Fragile Things: Short Fictions and Wonders*, London, Headline Publishing Group, 2006, p1.

⁴⁶⁶ P. Tabensky, 'The Self-Righteous Attack', *Mail and Guardian*, 29 September 2011, <http://mg.co.za/article/2011-09-29-the-selfrighteous-attack>.

which we must overcome. Rather, I have suggested that our vulnerability is something about our human situation which we must acknowledge and affirm.

Finally, it is important to point out that the arguments presented above should not be taken as evidence for the extreme, and implausible, view that our vulnerability is something which we should actively pursue as an end which is valuable in itself. As Nussbaum points out in her account of human vulnerability:

[T]his sensible reminder that a completely invulnerable life is likely to prove impoverished by no means entails that we should prefer risky lives to more stable lives, or seek to maximize our own vulnerability, as if it were a good in itself.⁴⁶⁷

The suggestion that we should strive to maximize our vulnerability, or see it as a valuable end in itself, is not only an extreme and implausible suggestion, but it would also be a particularly offensive view to defend. The vast majority of people across the world today live in extremely vulnerable situations – in many important ways not unlike the conditions of deprivation and confinement found in concentration camps. It is not just insensitive, but plainly offensive, to argue that such persons are living well because of their extreme vulnerability. Rather, my arguments in favour of the recognition and affirmation of our own vulnerability could provide reason for thinking that we should be more sensitive to the importance and impact of our human vulnerability on our lives, particularly when designing the political institutions which govern our society. To acknowledge and affirm our vulnerability might, in an important sense, also be to affirm our interconnectedness – to recognise one another as in a sense united by our shared vulnerability. Jointly affirming our shared vulnerability could, I think, form an important way to foster solidarity in human society, which remains divided along racial, ethnic, religious, gender and class lines. All of this, of course, is work for another occasion, but points to further avenues of thought to be pursued.

⁴⁶⁷ M. Nussbaum, *The Fragility of Goodness*, pp. xxix – xxx.

References

Apfel, L., *The Advent of Pluralism: Diversity and Conflict in the Age of Sophocles*, Oxford, Oxford University Press, 2011.

Arendt, H., *Eichmann in Jerusalem: a Report on the Banality of Evil*, London, Faber and Faber, 1963.

Ariely, D., and Loewenstein, G., 'When does Duration Matter in Judgment and Decision Making?', *Journal of Experimental Psychology: General*, vol. 129, 2000, pp. 508-29.

Ariely, D., and Zakay, D., 'A Timely Account of the Role of Duration in Decision Making', *Acta Psychologica*, vol. 108, no. 2, 2001, pp. 187-207.

Ariely, D., *Predictably Irrational: The Hidden Forces that Shape our Decisions*, New York, Harper, 2008.

Babcock, L., Loewenstein, G., and Issacharoff, S., 'Creating Convergence: Debiasing Biased Litigants', *Law & Social Inquiry*, vol. 22, no. 4, 1997, pp. 913-925.

Beauvoir, S., *The Ethics of Ambiguity*, Frechtman, B., trans., New York, Citadel Press, 1948.

Beauvoir, S., 'An Eye for an Eye', in Simons, M. (ed.), *Simone de Beauvoir: Philosophical Writings*, Urbana, University of Illinois Press, 2004.

Berlin, I., 'Two Concepts of Liberty', in Berlin, I., *Four Essays on Liberty*, Oxford, Oxford University Press, 1969.

Cooper, D., *Authenticity and Learning: Nietzsche's Educational Philosophy*, Great Britain, Routledge and Kegan Paul plc., 1991.

Cottingham, J., 'Integrity and Fragmentation', *Journal of Applied Philosophy*, vol. 27, no. 1, 2010, pp. 2-14.

Cox, D., La Caze, M., and Levine, M., *Integrity and the Fragile Self*, England, Ashgate Publishing Limited, 2003.

Danzigera, S., Levavb, J., and Avnaim-Pessoa, L., 'Extraneous Factors in Judicial Decisions', *Proceedings of the National Academy of Sciences of the United States of America*, Princeton, vol. 108, no. 17., 2011, pp. 6889-6892.

Davenport-Hines, R., *Gothic: 400 Years of Excess, Horror, Evil and Ruin*, London, Fourth Estate, 1998.

Day, W., *In the Circles of Fear and Desire: A Study of Gothic Fantasy*, Chicago and London, The University of Chicago Press, 1985.

Dennett, D., 'Conditions of personhood', in R. Rorty, (ed.), *The Identities of Persons*, Berkely, University of California Press, 1976, pp. 175-196.

Dworkin, G., 'Autonomy and Behaviour Control', *Hastings Center Report*, vol. 6, 1976, pp. 23-28.

Dworkin, G., *The Theory and Practice of Autonomy*, Cambridge, Cambridge University Press, 1988.

Elliott, A., *Psychoanalytic Theory: An Introduction, 2nd edn.*, Hampshire, Palgrave, 2002.

Elveton, R., 'Nietzsche's Stoicism: The Depths are Inside', in Bishop, P., (ed.), *Nietzsche and Antiquity*, Rochester, Camden House, 2004, pp. 192-203.

Emerson, R., 'Self-Reliance', in *Essays and Lectures*, New York, Library of America, 1983.

Frankfurt, H., 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, vol. 68, no. 1, 1971, pp. 5-20.

Frankfurt, H., 'Identification and wholeheartedness', in Frankfurt, H., *The Importance of what we Care About*, Cambridge, Cambridge University Press, 1988.

Frankfurt, H., 'The Faintest Passion', in Frankfurt, H., *Necessity, Volition and Love*, Cambridge, Cambridge University Press, 1999.

Frankfurt, H., 'Autonomy, Necessity, and Love', in Frankfurt, H., *Necessity, Volition and Love*, Cambridge, Cambridge University Press, 1999.

Frankfurt, H., 'Taking Ourselves Seriously', in Satz, D. (ed.), *Taking Ourselves Seriously & Getting It Right*, California, Stanford University Press, 2006.

Frankfurt, H., 'Reply to J. David Velleman', in Buss, S., and Overton, L., (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt*, United States of America, MIT Press, 2002, pp. 124-128.

Frankl, V., *Man's Search for Meaning*, Ilse Lasche, I. trans., Boston, Massachusetts, Beacon Press, 2006.

Freeman, M., 'Death, Narrative Integrity, and the Radical Challenge of Self-Understanding: a Reading of Tolstoy's *Death of Ivan Ilych*', *Ageing and Society*, vol. 17, no. 4, 1997, pp. 373-398.

Friedman, M., 'Autonomy and the Split-Level Self', *The Southern Journal of Philosophy*, vol. XXIV, no. 1., 1986, pp. 19-35.

Friedman, M., 'The Impracticality of Impartiality', *The Journal of Philosophy*, vol. 86, no. 11, 1989, pp. 645-656.

Freud, S., *An Outline of Psychoanalysis*, Strachey, J. trans., London, The Hogarth Press and the Institute of Psycho-analysis, 1949.

Freud, S., *Introductory Lectures on Psycho-Analysis*, Riviere, J. trans., New York, Columbia University Press, 1983.

Gailliot, M., Baumesiter, R., DeWall, N., Maner, J., Plant, E., Tice, D., Brewer, L., and Schmeichel, B., 'Self-Control Relies on Glucose as a Limited Energy Source: Willpower Is More Than a Metaphor', *Journal of Personality and Social Psychology*, vol. 92, no 2., 2007, pp. 325-336.

Gaiman, N. *Fragile Things: Short Fictions and Wonders*, London, Headline Publishing Group, 2006.

Garrett, P., *Gothic Reflections: Narrative Force in Nineteenth-Century Fiction*, London, Cornell University Press, 2003.

Gemes, K., and May, S., (eds.), *Nietzsche on Freedom and Autonomy*, Oxford, Oxford University Press, 2009.

Gemes, K., 'Nietzsche on Free Will, Autonomy and the Sovereign Individual', in Gemes, K., and May, S., (eds.), *Nietzsche on Freedom and Autonomy*, Oxford, Oxford University Press, 2009.

Gert, J., *Brute Rationality: Normativity and Human Action*, Cambridge, Cambridge University Press, 2004.

Gilbert, D. T., and Jones, E. E., 'Perceiver-Induced Constraint: Interpretations of Self-Generated Reality', *Journal of Personality and Social Psychology*, vol. 50, 1986, pp. 269-280.

Guignon, C., *On Being Authentic*, London and New York, Routledge, 2004.

Haney, C., Banks, C., and Zimbardo, P., 'A Study of Prisoners and Guards in a Simulated Prison', *Naval Research Reviews*, 9, Washington, Office of Naval Research, 1973, pp. 1-17.

Halberstam, J., *Skin Shows: Gothic Horror and the Technology of Monsters*, Durham and London, Duke University Press, 1995.

Heidegger, M., *Nietzsche; Volume 2: The Eternal Recurrence of the Same*, Krell, D., trans., New York, Harper & Row Publishers, 1984.

Heinämaa, S., *Toward a Phenomenology of Sexual Difference*, United States of America, Rowman and Littlefield Publishers, Inc., 2003.

Holland, A., *Excess and Transgression in Simone de Beauvoir's Fiction: The Discourse of Madness*, Farnham, Ashgate Publishing Ltd., 2009.

Horsthouse, R., *On Virtue Ethics*, New York, Oxford University Press Inc., 1999.

Hume, R. D., 'Gothic vs Romantic: A Revaluation of the Gothic Novel', *PMLA*, vol. 84, no. 2, 1969, pp. 282-290.

Jackson, R., *Fantasy: The Literature of Subversion*, London, Methuen, 1981.

Jones, D., *Horror: A Thematic History in Fiction and Film*, London, Arnold, 2002.

Lear, J., *Open Minded: Working Out the Logic of the Soul*, Cambridge, Massachusetts, Harvard University Press, 1998.

L'Engle, M., *Walking on Water: Reflections on Faith and Art*, Farrar, Straus and Giroux, 1995.

Lessing, A., 'What is Wrong with a Forgery?', in Warburton, N., (ed.), *Philosophy: Basic Readings*, 2nd Edition, London and New York, Taylor and Francis, 2005, pp. 536-549.

Loewenstein, G., 'Out of Control: Visceral Influences on Behavior', *Organizational Behaviour and Human Decision Processes*, vol. 65, no. 3, 1996, pp. 272-292.

Loewenstein, G., 'Willpower: A Decision-Theorist's Perspective', *Law and Philosophy*, vol. 19, 2000, pp. 51-76.

Martel, Y., *Life of Pi*, New York, Harcourt Inc., 2001.

Massé, M. 'Psychoanalysis and the Gothic', in Punter, D., (ed.), *A Companion to the Gothic*, United States of America, Blackwell, 2001, pp. 229-241.

May, S., 'Nihilism and the Free Self', in Gemes, K., and May, S., (eds.), *Nietzsche on Freedom and Autonomy*, Oxford, Oxford University Press, 2009.

Mele, A., *Irrationality: An Essay on Akrasia, Self-deception and Self-control*, Oxford, Oxford University Press, 1987.

Mele, A., *Autonomous Agents: From Self-Control to Autonomy*, New York, Oxford University Press, 1995.

Mele, A., *Motivation and Agency*, Oxford, Oxford University Press, 2003.

Mele, A., *Backsliding: Understanding Weakness of Will*, Oxford, Oxford University Press, 2012.

Mele, A., 'Autonomy and Neuroscience', in Radoilska, L. (ed.), *Autonomy and Mental Disorder*, Oxford, Oxford University Press, 2012.

- Midgley, M., *Beast and Man: The Roots of Human Nature*, London, Methuen, 1979.
- Midgley, M., *The Ethical Primate: Humans, Freedom, Morality*, London, Routledge, 1994.
- Midgley, M., *Wickedness: A Philosophical Essay*, London, Routledge, 2000.
- Milgram, S., 'Behavioural Study of Obedience', *The Journal of Abnormal and Social Psychology*, vol. 67, no. 4, 1963, pp. 371-378.
- Mill, J., *On Liberty*, Harmondsworth, Penguin, 1985.
- Miyoshi, M., *The Divided Self: A Perspective on the Literature of the Victorians*, New York, New York University Press, 1969.
- Moran, R., *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton, Princeton University Press, 2001.
- Murphy, N., and W. Brown, *Did My Neurons Make Me Do It? - Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will*, Oxford, Oxford University Press, 2007.
- My Way, P. Anka, writer, F. Sinatra, performer, Reprise, Los Angeles, 1968.
- Napier, E., *The Failure of the Gothic: Problems of Disjunction in an Eighteenth Century Literary Form*, Oxford, Oxford University Press, 1987.
- Nehamas, A., 'The Eternal Recurrence', in Richardson, J., and Leiter, B., (eds.), *Nietzsche*, Oxford, Oxford University Press, 2001, pp. 120 – 138.
- Nehamas, A. 'How One Becomes What One Is', in Richardson, J., and Leiter, B., (eds.), *Nietzsche*, Oxford, Oxford University Press, 2001, pp. 255-280.
- Nietzsche, F., *The Birth of Tragedy from the Spirit of Music*, Kaufmann, W., trans., *The Basic Writings of Nietzsche*, New York, Toronto, The Modern Library, 2000.
- Nietzsche, F., *Beyond Good and Evil: Prelude to a Philosophy of the Future*, Norman, J., trans., UK and New York, Cambridge University Press, 2002.
- Nietzsche, F., *The Will to Power*; Kaufmann, W., and Hollingdale, R. J., trans., New York, Vantage Books, 1967.
- Nietzsche, F., *On the Genealogy of Morals*, Kaufmann, W., trans., *The Basic Writings of Nietzsche*, New York, Toronto, The Modern Library, 2000.
- Nietzsche, F., *Twilight of the Idols; The Anti-Christ*; Hollingdale, J. trans., Radice, B., and Baldick, R., (eds.), Great Britain, Penguin Books Ltd., 1969.

Nietzsche, F., *The Gay Science*, Nauckhoff, J., trans., Cambridge, Cambridge University Press, 2001, pp. 194 - 195.

Nietzsche, F., *Thus Spoke Zarathustra*, in Hollingdale, R., (ed.), *A Nietzsche Reader*, London, Penguin, 1977.

Nietzsche, F., *Schopenhauer as Educator*, in Hollingdale, R., trans., *Friedrich Nietzsche: Untimely Meditations*, Cambridge, Cambridge University Press, 1989.

Noggle, R., 'Autonomy and the Paradox of Self-Creation: Infinite Regresses, Finite Selves, and the Limits of Authenticity', in J. Taylor, (ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, Cambridge, Cambridge University Press, 2005, pp. 87-108.

Nussbaum, M., 'Mortal Immortals: Lucretius on Death and the Voice of Nature', *Philosophy and Phenomenological Research*, vol. 50, no. 2, 1989, pp. 303-351.

Nussbaum, M., 'Pity and Mercy: Nietzsche's Stoicism', in Schacht R. (ed.) *Nietzsche, Genealogy, Morality*, Berkeley, California, University of California Press, 1994, pp. 139-167.

Nussbaum, M., *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*, 2nd edn., Cambridge, New York, Cambridge University Press, 2001.

Paphitis, S., 'Questions of the Self in the Personal Autonomy Debate: Some Critical Remarks on Frankfurt and Watson', *The South African Journal of Philosophy*, vol. 29, No. 2, 2010, pp. 57-71.

Pippin, R., 'How to Overcome Oneself: Nietzsche on Freedom', in Gemes, K. and May, S., (eds.) *Nietzsche on Freedom and Autonomy*, Oxford, Oxford University Press, 2009, pp. 69-88.

Rank, O., *The Double: A Psychoanalytic Study*, Tucker, H., trans., (ed.), United States of America, The University of North Carolina Press, 1971.

Reginster, B., 'Nietzsche on *Ressentiment* and Valuation', *Philosophy and Phenomenological Research*, vol. LVII, no. 2, 1997, pp. 281-305.

Rorty, A., 'How to Harden your Heart: Six Easy Ways to become Corrupt', in Rorty, A., (ed.), *The Many Faces of Evil: Historical Perspectives*, London and New York, Routledge, 2005.

Ross, L., and Nisbett, R., *The Person and the Situation: Perspectives of Social Psychology*, New York, McGraw-Hill, 1991.

Ruddick, W., and Rachels, J., 'Lives and Liberty', in Christman, J., (ed.), *The Inner Citadel: Essays on Individual Autonomy*, Oxford, Oxford University Press, 1989, pp. 221-233.

Sandler, J., and Wallerstein, R., *Freud's Models of the Mind: An Introduction*, Madison, International Universities Press, 1998.

Shakespeare, W., *Hamlet*, United Kingdom, Penguin Classics, 1994.

Stevenson, R., *The Strange Case of Dr Jekyll and Mr Hyde*, Danahay, M., (ed.), Canada, Broadview Press Ltd., 2000.

Styron, W., *Sophie's Choice*, New York, Random House, 1976.

Tabensky, P., 'The Self-Righteous Attack', *Mail and Guardian*, 29 September 2011, <http://mg.co.za/article/2011-09-29-the-selfrighteous-attack>.

Taylor, C., *Sources of the Self: The Making of the Modern Identity*, Cambridge, Cambridge University Press, 1989.

The Manchurian Candidate, dir. Frankenheimer, J., USA, M. C. Productions, 1962.

Thompson, G., *The Gothic Imagination: Essays in Dark Romanticism*, Pullman, Washington State University Press, 1974.

Tversky, A., and Kahnemann, D., 'Judgment under Uncertainty: Heuristics and Biases', *Science, New Series*, vol. 185, no 4157, 1974, pp. 1124-1131.

Velleman, J., *Self to Self: Selected Essays*, Cambridge, Cambridge University Press, 2006.

Velleman, J., 'Identification and Identity', in Buss, S., and Overton, L., (eds.), *Contours of agency: Essays on Themes from Harry Frankfurt*, United States of America, MIT Press, 2002, pp. 91-123.

Wallace, R., *The View from Here: On Affirmation, Attachment, and the Limits of Regret*, Oxford, Oxford University Press, 2013.

Watson, G., 'Free Agency', in Christman, J. (ed.), *The Inner Citadel: Essays on Individual Autonomy*, Oxford, Oxford University Press, 1989.

Watson, G., *Agency and Answerability: Selected Essays*, New York, Oxford University Press, 2004.

Wilson, T., and Brekke, N., 'Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations', *Psychological Bulletin*, vol. 116, no. 1., 1994, pp. 117-142.

Wolf, S., 'Sanity and the Metaphysics of Responsibility', in Schoeman, F. (ed.), *Responsibility, Character and the Emotions: New Essays in Moral Psychology*, Cambridge University Press, 1998.

Wolfreys, J., 'I Wants to Make Your Flesh Creep', in J. Wolfreys, and R. Robbins, (eds.), *Victorian Gothic: Literary and Cultural Manifestations in the Nineteenth Century*, Basingstoke, Palgrave Macmillan, 2000, pp. 31-59.

Zimbardo, P., and Haney, G., 'The Past and Future of U.S. Prison Policy Twenty-Five Years After the Stanford Prison Experiment', *American Psychologist*, vol. 53, no. 7, 1998, pp. 709-727.

Zimbardo, P., *The Lucifer Effect: Understanding How Good People Turn Evil*, New York, Random House, 2007.