Towards a Horizontal Theory of Autonomy

Fiona Whittingham

UCL

MPhil Stud Philosophy

I, Fiona Whittingham, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The thesis criticises 'hierarchical' accounts of personal autonomy, and outlines the core features of an alternative, 'horizontal' account. It specifically focuses on the conditions of autonomous action.

The current dominant approaches to the phenomenon of personal autonomy, or self-governance, tend to isolate certain privileged aspects of the self, that wield control when an agent is autonomous. In this sense, they can be labelled as 'hierarchical' accounts. For instance, Frankfurt's early (1971) account bases the conditions for autonomous action in conformity with an individual's higher-order volitions. However, I argue that such 'hierarchical' approaches are unsatisfactory. Bringing in considerations from Buss (1994, 2013), I argue that they are unable to adequately account for the phenomenon of nonautonomous action, or to explain how these privileged aspects of the soul exert authority over the rest.

Instead, I argue for a 'horizontal' account of the self-relation involved when an individual acts autonomously. Such an account does not locate the conditions for autonomous action in some limited subsection of the self, but instead takes a more holistic approach to the aspects of the self that can be involved in self-governance. I suggest that the positive conditions for personal autonomy can instead be understood to involve the individual's conceptualisation of the decision-situation before them, and the range of factors incorporated into their decision-making. Under this kind of account, personal autonomy is a notion admitting of degrees.

Impact Statement

This thesis criticises the dominant structure adopted by contemporary accounts of personal autonomy, which I label a 'hierarchical' structure, and makes the case for a new, horizontal alternative. It points towards a new way of understanding the self-relation involved in autonomy, that can provide an alternative object of investigation to existing accounts. It also suggests ways of explaining low autonomy versus high autonomy states that could have practical application outside academia, for instance in the treatment of addiction.

Contents

Tables and Figures

Towards a Horizontal Theory of Autonomy

Chapter 1: Introduction

Introduction

In this thesis, I will be arguing that we should be moving towards a 'horizontal' account of autonomy. I will criticise dominant, 'hierarchical' accounts (chapter two) and make the case for moving towards a 'horizontal' alternative (chapters three and four).

In this chapter I will introduce the concept that will be the topic of this thesis: that of personal autonomy. I will run through a number of ways of construing the concept, and present some examples of failures of autonomy. This will lead to an assessment of a set of core features or requirements that an adequate account of personal autonomy should incorporate. I will then outline the plan for the following chapters, and the aims for the thesis as a whole.

Personal Autonomy

Personal autonomy can be understood as a capacity to lead a self-directed, or a self-governed, life. When a person acts autonomously, their action issues from themselves in a deeper and a truer sense than when they fail to act autonomously. The values, commitments, or desires that are embodied in their action can be said to represent themselves, and accordingly their action is their own. The question of what it means to be autonomous accordingly links to the question of what the self is, or which are the core or central aspects of the self.

One way of accessing the concept of personal autonomy is through an analogy with a political governance relation (Buss 2018). When a government is authoritative, it has the power to enact laws in the relevant country. A governance relation holds between the government and the citizens of the country. The government constrains what it is that its people can (legitimately, or at least legally) do. Autonomy can be understood as self-governance. When an agent self-governs, it sets limits to what the self can do, as long as it is to yield to the authority of the governing self. The agent is autonomous when this authority is adhered to. However, the disanalogy with the legal case is that the one who governs and the one who is governed are contained within the confines of a single individual. This adds

complexity, since there is no simple way of distinguishing the self that governs from the self that is governed.

Since the self that governs and the self that is governed are numerically one and the same individual, it seems hard to understand how it can be possible for self-governance to *fail* to occur. Why would an agent act against its own dictates? And if it does, how can the resulting behaviour be ascribed to the agent as its action at all? This is the problem of understanding the possibility of nonautonomous action, which I will discuss in the next chapter. Moreover, given that the two sides of the self-relation can come apart in this way, there is the problem of isolating the mechanism by which the governing side of the self-relation wields authority when the conditions for self-governance are fulfilled. How does the self-governing agent wield authority over itself? Once again, this is harder to conceptualise in the case in which the one who governs and the one who is governed are within a single individual, since the mechanism seemingly must be internal, and it might not necessarily be overtly displayed. This again I will discuss in later chapters.

Another way of conceptualising the problem of autonomy is that of distinguishing between those influences upon our actions and behaviour that are autonomy-undermining, and those that are autonomy-conferring (Buss 2018). As creatures of the world, everything we do can be ascribed to origins that we have no control over, whether as the result of random factors or otherwise. Therefore, the issue when it comes to nonautonomous action cannot be that our actions have their origins in factors external to ourselves. In that case, we could never be autonomous. Rather, the issue for personal autonomy is distinguishing between the case in which a successfully governing agent emerges out of this series of external events, and one in which this never occurs, such that the agent is simply at the whim of their resulting actions or behaviour.

Examples of Failures of Autonomy

The explanandum of this thesis can also be targeted by elucidating instances of its failure: cases in which agents fail to be autonomous, or to act autonomously. A classic case is the case of addiction. When an individual is addicted to some kind of substance, or activity, they are intuitively moved to act by forces that are not their own. They do not have the kind of control or governance over their behaviour that is required for autonomy. For instance, a drug addict might carry out acts of consumption of the specific drug, and actions in the process of

getting hold of the drug, that do not seem to be their own. Let us say that they truly lament their drug habit, and wish to give up, but despite their intentions and resolutions they continue to relapse. They may also act in ways that are out of character in order to get hold of the drug, such as stealing. Furthermore, their actions might go against their deepest commitments and concerns, for example they take the drug instead of feeding their child dinner. All of this suggests that their actions do not reflect their true self, or that there is a failure in the authority relation that would normally hold between them and their actions.

I will later return to the addiction example in the context of Heyman's observations about addiction in his *Addiction as Choice* (2010). Whilst addiction seems to involve an autonomy-undermining influence upon behaviour, that subverts the decision-making process, other examples of autonomy-deprivation involve cases in which individuals do not have adequate resources to develop an autonomous intention or attitude. Perhaps they might lack the kinds of reflective and rational abilities that are the usual possession of adult humans. Frankfurt (1971) argues that it is constitutive of personhood to take attitudes towards our own motivational states. Whether this is the case or not, if an individual is lacking in this kind of human reflexivity, or alternatively lacks the ability to recognise and respond to reasons, then we are unlikely to label them as an autonomous agent. Examples may include individuals with severe mental disability. It also seems that an individual can be lacking in autonomy when they do not possess an adequate range of action options to choose between, or do not have access to an adequately rich normative framework within which to make decisions and build their lives. I will discuss a detailed example of this kind of autonomy-deprivation in chapter three, concerning the Crow nation of north America and the cultural devastation they underwent following the encroachment of European settlers (from Lear, 2008).

Requirements for an Adequate Theory of Personal Autonomy

Altogether, an adequate theory of personal autonomy should find a way of explaining how a self-governing agent can emerge out of a series of prior events and influences, none of which are their own doing. It should explain how nonautonomous action can be possible, and should also explain the mechanism by which successful self-governance can be wielded. In doing so, it should help explain why autonomous actions can more truly be described as the agent's own, and why certain actions but not others are representative of the true self.

The kinds of examples of autonomy-deprivation an adequate theory of personal autonomy should be able to explain include the case of addiction, together with inadequate rational and reflexive capacities. An adequate account of personal autonomy should also be able to account for the kind of autonomy deprivation that can occur when an individual is lacking in opportunities in terms of action options to choose from, or an adequately rich normative framework.

Plan for the Rest of the Thesis

Over the rest of the thesis, I will argue that the dominant form of approach to explaining personal autonomy – the 'hierarchical' approach – is inadequate. It fails to explain the possibility of nonautonomous action, or the mechanism by which the authority relation involved in personal autonomy can be wielded. Instead, I will argue that we should be moving towards a 'horizontal' alternative.

In the next chapter, chapter two, I will set out my critique of the dominant 'hierarchical' approaches to personal autonomy. I will carry out an overview of these kinds of approaches, and discuss the features they share. I will then detail two key issues with such accounts: (a) explaining nonautonomous action, and (b) integration.

In chapter three, I will outline the shape of a horizontal alternative. I will discuss the way in which I propose we level the hierarchy implied by hierarchical accounts, and move towards a more holistic view, which considers the self as a whole when it comes to the exertion of authority involved in personal autonomy. I will consider the intuitive pull of higher-order mental states and 'all-things-considered' judgements when it comes to autonomy, and specific examples of autonomy-deprivation. Emerging from this, I will argue for a 'horizontal' view according to which autonomy is tied to the richness and breadth of the perspective adopted by an agent. I will set out a range of options for ways in which the detail of the horizontal account could be filled in, though the precise shape a horizontal account would best take is beyond the scope of this thesis.

In chapter four, I will discuss the ways in which the horizontal account is able to bypass the issues with hierarchical accounts discussed in chapter one. I will also set out a positive argument in favour of it, in terms of the nature of practical deliberation and observations that can be made about practical deliberation from a third-person perspective.

I will set out my conclusions in chapter 5.

Chapter 2: Critique of 'Hierarchical' Approaches to Personal Autonomy

Introduction

In this section, I will set out my critique of the dominant type of account of personal autonomy – 'hierarchical' accounts. First, I will carry out an overview of some 'hierarchical' approaches to personal autonomy, and the reasoning behind them. I will highlight the key features they share that will be the target of my criticism. Drawing on arguments from Sarah Buss (1994, 2013), I will then argue that they fail on two fronts: they fail to explain the possibility of nonautonomous action, and they fail to explain the kind of integration required for an agent to exert authority over themselves. For these reasons, I propose moving towards a 'horizontal' alternative, the outlines of which I will set out in the next chapter.

'Hierarchical' Accounts of Personal Autonomy

Most of the dominant approaches to personal autonomy follow a type of structure that I will in this thesis label as 'hierarchical'. 'Hierarchical' approaches answer the questions around personal autonomy set out in the previous chapter by isolating certain privileged aspects of the self that exert control when an agent self-governs, for instance an agent's higher-order desires (Frankfurt 1971), or their judgement about the best thing to do (e.g. Moran (2002), Raz (2002)). The relevant core aspects of the self represent the 'true' self, which helps to explain why actions that align with those aspects of the self are owned and self-directed in a sense in which other, nonautonomous actions are not. An agent is autonomous when these core aspects of the self successfully self-govern. Nonautonomous action occurs when there is a misalignment between these core aspects and the motivational forces that successfully lead an individual to action. Autonomy-undermining influences upon action are those that lead an individual to act in such a misaligned way, creating conflict between the way in which an individual actually behaves, and their core, 'true' self.

For (early) Frankfurt (1971), the thing that distinguishes persons is their capacity to form volitions of the second order. Second order volitions are a species of second order desire – desires that an agent has about their desires. Humans are distinguished from the animals in their capacity to form such second order desires – desires to have (or not to have) certain desires and motives. However, for Frankfurt, what is integral to the concept of a *person* is

their capacity to take a stance in relation to which of their first order desires they would like not only to have, but also to be effective in leading them to action. These are their second order *volitions* on what they would like the shape of their will to be. To be a person, an agent needs to care about the direction their will goes in. Otherwise, they will be what Frankfurt labels a mere 'wanton' (p11), without concern for which motives lead them to behave in one way or another. The capacity to form second-order volitions depends on an agent's rational capacities, but also involves more, since an agent could be instrumentally rational in relation to their desires, but not care about which desires are ultimately the ones that move them to behave in one way or another.

For Frankfurt, freedom of the will requires that an agent is free to have the will that he wants – he is free in that his second-order volitions determine the way in which his will in fact goes. Freedom of action, on the other hand, is the freedom to *do* what it is that one wants to do. It means that one's desires are satisfied in action. For Frankfurt, if an individual possesses both freedom of the will and freedom of action, he therefore has (p18):

> 'all the freedom it is possible to desire or conceive'

If an individual possesses freedom of the will and freedom of action, then not only is he free to determine the way in which his will goes, but he will also find that his will is satisfied when it comes to action. Frankfurt's account suggests that the relevant corresponding actions will be autonomous. We act autonomously when we act on the basis of desires which we endorse. Frankfurt goes on to grant the possibility of higher-order volitions still – third or fourth order desires that certain first order desires be the ones to lead an individual to action. However, when an individual forms a *decisive* higher-order commitment about the shape they would like their will to be, this sets in place the relevant benchmark for determining the freedom of the will.

Altogether then, for (early) Frankfurt, our higher-order volitions – our desires about which first-order desires are to be effective in moving us to action – represent the true self, and when we act in accordance with these higher-order volitions, we act autonomously. Self-governance involves our higher-order mental states wielding authority over the motives that are manifested in action. The human capacity for reflexivity – to take a stance with respect to aspects of our mental lives – is what makes this kind of relation to aspects of ourselves possible. Therefore, Frankfurt takes the aspect of the self that he deems integral to the concept of a person – their set of higher-order volitions – and treats this as the benchmark for

determining the conditions of autonomous action. The decisive higher-order volitions are the authoritative aspects of the self, that determine whether an action that results from an agent's motivational profile can truly be considered their own. Nonautonomous action occurs when motivational forces other than those sanctioned by an agent's decisive higher-order volitions lead an agent to behave in a certain way. Frankfurt presents as an example the 'unwilling addict' (p12). He has a desire both to take the drug, and not to take the drug, but his second-order volition is that the desire not to take the drug is the one that is effective in determining how he behaves. However, he is 'helplessly violated' by his first-order desire to take the drug (p12). In taking the drug, he does something he indeed wants to do, and in this sense he has freedom of action. However, by identifying with one desire rather than the other he makes the former 'more truly his own' (p13). Therefore, we can say that the motivational underpinnings of his behaviour are not his own, despite it also being in a sense his own free action.

Other hierarchical accounts treat as the authoritative aspects of the self a set of core commitments or values, which may or may not be said to compose an agent's 'character'. For instance, Watson (2004) argues that Frankfurt's (1971) higher-order desires account leads to regress problems, which Frankfurt's notion of a 'decisive commitment' is inadequate to resolve. In response, he argues for distinguishing between a person's valuational system, which includes their long term ends and normative principles, and their motivational system, that involves their desires with their various orders and strengths, the weighting of which will determine how they in fact act. Unfreedom or alienation is possible under Watson's account when a person's valuational and motivational systems misalign, for example when a person is driven to pursue short-term pleasures that they do not care about at the expense of neglecting the things they really value, like their career and their family. For Watson then, a person's valuational system is the relevant authoritative aspect of the self that determines the conditions of autonomy.

In a similar vein, in later work Frankfurt (2002, 2006) shifts away from his hierarchical desires-based approach to delineating the motivational states that an agent identifies with, to one based around the idea that identification is a *reason-generating* volitional state (2002, reply to Moran). For later Frankfurt, when we identify with an element of our psychology, the paradigm case being the love we hold for someone (2002), that mental state provides us with binding reasons for action, and sets limits on what it is possible for us to will. When we love someone, we cannot help but be motivated to pursue their best interests, and when we are in the grip of that state, we cannot help but love them. Therefore, autonomy is interestingly tied

to restrictions on our possible choices. Love constrains what it is possible for us to will. Perhaps we could attempt to manipulate our love, but we would have to do so in indirect ways like the ways in which we might attempt to manipulate the mental states and attitudes of others. This notion of volitional necessities is used to isolate those things we identify with, i.e. those things that we care about or our core commitments. These commitments form the privileged aspects of our motivational profile, and when we act in alignment with them, we are autonomous.

An example of a character-based account is Ekstrom's (1993), which applies a coherentist approach to such a structure. Ekstrom (1993) argues for an alternative account of autonomy and the self to Frankfurt's early (1971) account, according to which the self is to be construed as one's system of acceptances and preferences. Acceptances pertain to an individual's set of beliefs, whilst preferences are a subspecies of higher-order desire. Preferences are similar to Frankfurt's (1971) higher-order volitions, but they fulfil the additional condition of having been formed in the search for the good. Since it is an evaluative faculty that performs this role, as opposed to simply a higher-order desire that could be based on anything, it is meant to be more convincing that preferences, as opposed to mere higher-order volitions, can be regarded as internal to the self without recourse to a regress of further higher-order mental states. They are meant to provide a plausible grounding for the integral aspects of the self. Autonomous action is specifically described as behaviour that issues from an agent's authorised preferences, with an agent's authorised preferences being those that cohere with the other preferences and acceptances in the individual's character system. Once again, then, Ekstrom (1993) defines autonomous action via its alignment with a set of core commitments or values. An individual fails to be autonomous when they are moved to act by influences that fall outside of this core, and thus that run contrary to the dictates of their 'true self'.

Still other 'hierarchical' accounts treat as the core aspect of the self an individual's rational nature, or their best judgement about what to do in a situation (or some more sophisticated variation on the notion of a best judgement). For Raz (2002), Frankfurt's distinction between the active and the passive (as Frankfurt later construes it) is not to be explicated by some form of endorsement of the relevant mental states or events, but rather by the extent to which the relevant mental states or events are responsive to reasons, as the agent understands them. Raz goes on to explain (p16-17):

'We believe that we are properly responsive to reason whenever we would, if challenged, resist, at least initially, claims that our conduct, intentions, emotions, etc., are irrational or unreasonable'

Raz begins with the case of the belief, in his attempt to find the marker of the active as opposed to the passive in mental life. He notes that when it comes to belief, we tend to consider ourselves to be active, despite beliefs not being subject to choice. Rather, our beliefs are mainly active because they adjust according to the evidence available – they are reasons-responsive. The account is then extended to other aspects of our mental lives, like action and desire. Raz goes on to stipulate that the relevant responsiveness to reasons is responsiveness to reasons as we understand them, which helps to explain why we can be active but irrational, for instance in cases of akrasia in which we provide to ourselves rationalising explanations of why our behaviour is correct. Like Frankfurt, Raz supports his account with an appeal to personhood, since the relevant form of activity is associated with life, as opposed to passivity (and death) (p211). This helps in supporting the contention that a person's rational nature is the relevant core aspect of the self, which determines the conditions for autonomy. For Raz, then, the aspect of the self which forms the top of the hierarchy is an agent's rational nature, or their judgement about what is best or right. This determines which mental states and actions are their own, as opposed to being processes or happenings that simply befall them.

There are other accounts of personal autonomy which offer variation on the notion of an agent's best judgement. For instance, Denis McManus (2019) argues for an interpretation of Hiedegger's (SZ) notion of authenticity or 'ownness' according to which authentic action is the manifestation of 'all-things-considered judgements' (p1194). McManus notes that in any decision-situation in which an individual finds themselves, they will be subject to multiple competing normative demands, for instance they will possess certain roles, and be in the pursuit of multiple goals and projects. Perhaps we might add that they face certain moral normative demands simply in virtue of being a member of the human race. When an individual makes an all-things-considered judgement, they take into account the entire situation as it presents itself to them, including all its normative pulls in one direction or another, and accordingly decide how to act. They make a judgement about what the situation calls for in its entirety, recognising that in their decision to promote certain of their aims and purse certain demands of the situation, they will have neglected certain other aims and demands.

This kind of all-things-considered judgement is the model of authentic action, to be contrasted with action that reflects what McManus terms a mere 'Q judgement' (p1194) – a judgement carried out from the point of view of some particular goal, or some particular role. A Q judgement assesses what the situation calls for from a limited, qualified perspective, and if an individual decides how to act in this way, they may be understood as attempting to absolve themselves of responsibility for their actions, passing this on to some set of normative demands or another. In contrast, in an all-things-considered judgement, they take responsibility for choosing between the multiple normative demands of the situation, thus expressing *themselves* as opposed to just one particular normative standpoint. Therefore, for McManus, the core, authoritative self is the enactor of all-things-considered judgements, or the self who is able to step aside from mere particular Q judgements to take a stance on what a situation calls for as a whole.

Criticism (1): Explaining the Possibility of Nonautonomous Action

Each of these 'hierarchical' accounts, therefore, has a story about what goes on when an agent acts autonomously. However, it is not so clear that they can adequately explain what goes on when an agent acts in a way that is not autonomous. When it comes to autonomy, we still want to explain cases in which an agent fails to successfully self-govern in their actions. In such cases, they are moved by forces that are not fully their own, but that still fulfil the minimal conditions required in order for us to label the relevant happenings as their action, and even their intentional or goal-directed action. The concept of action is used to isolate the movements of the body, or other changes in a person, that can be ascribed to the person as a whole. The label of 'intentional action', alternatively, tends to be reserved for action that is goal-directed. Autonomous action, on the other hand, fulfils further requirements in order for such actions to be self-directed, and seems to be a distinct category from mere intentional action, or action in general.

That an agent can act, and even intentionally act, in ways that are not autonomous is indicated by the examples set out in the previous chapter. An addict might intentionally carry out many actions that we do not consider to be their own in the sense required for us to label their actions as autonomous. They might exhibit instrumental rationality in the pursuit of certain goals, such as, in an extreme case, stealing in order to get hold of money required for a drug. However, as they are moved by the force of their desire for the drug in this way, and contrary

to things they care about like their work and their family, we do not understand their intentional behaviour to be autonomous. Alternatively, there might be individuals, or even animals, who display goal-directed behaviour, but that do not possess sophisticated reflective or rational capacities, and we do not consider them to fulfil the requirements for autonomy. Further still, individuals with a minimal set of options to choose from may convincingly be understood to act and pursue goals, but we may not consider them to be doing so autonomously.

For 'hierarchical' accounts, an agent acts autonomously when their actions align with some dominant aspect of the self that represents the 'true' or authoritative self, such as an agent's set of decisive higher-order volitions (Frankfurt 1971), core commitments that might be said to compose a 'character' (Frankfurt 2002, Watson 2004, Ekstrom 1993), or some variation on the notion of a best judgement about what to do in a situation (Raz 2002, McManus 2019). However, it is less clear why this 'dominant' or true self would allow itself to be superseded by forces that are not its own. If it is indeed overcome in such a way, then it is not so clear that the resulting behaviour can be convincingly ascribed to the agent as a whole, as its intentional action, or even simply as its action. The authoritative, or official and sanctioned, motivational aspects of the self, that, when successfully yielded to, represent the self in action, are superseded. The relevant forces that they are superseded by, whilst being in an empirical sense internal to the self, are not considered to be the self's own. It is not so clear that the resulting behaviour is an action that can be ascribed to the agent as a whole, at all. According to this characterisation of 'nonautonomous action', it is as though the self has been sequestered by some foreign power.

Buss (1994) provides an objection to endorsement-type accounts of personal autonomy that can be applied as a general criticism of the way in which 'hierarchical' accounts attempt to delineate autonomous intentional action from nonautonomous intentional action. Buss argues that the notion of endorsement cannot be used to secure the conditions for autonomy, since if an individual acts contrary to their all-things-considered preference, then they are not acting intentionally at all. Therefore, for Buss, the notion of endorsement or identification, as broadly construed to incorporate variations that appeal to higher-order desires, but also variations that appeal to considered values and abilities to endorse, cannot be used to explicate autonomous action, as a distinct category from intentional or goal-directed action in general. Buss argues that when it comes to intentional or goal-directed action, an individual would not set herself the following kind of goal (p97):

'the goal of doing what she prefers *not* to do, all things considered'

Preferences, as opposed to mere inclinations, result from an agent's comparative evaluations about the merits of one course of action versus another. The idea is that when it comes to a specific action, and the situation as it presents itself to an individual, if an individual determines that they prefer not to do something, then it shows that they take themselves to have insufficient reason to do that thing, in comparison to the other options available. Therefore, they will not set themselves the goal of doing it, so it cannot be that they carry it out in an intentional manner. If they do implement such behaviour, it will conflict with their assessment or reasoned judgement of what the situation demands, and so it cannot be considered their own intentional action at all. Similarly, in later work (2013), Buss argues that the human capacity for reflexivity – to take a stance in relation to the motivational underpinnings of our behaviour – opens the potential for us to distance ourselves from the psychic sources of our behaviour in such a way that that behaviour might no longer qualify as our action. If we are moved by a force which we reject, then it is not clear that the resulting behaviour can truly be considered to be our action.

Take the unwilling addict. For Buss (1994), this example, as often construed, in fact presents a conceptual impossibility. It is meant to be that the unwilling addict intentionally, but nonautonomously, injects themselves with the drug, counter to the behaviour that they, as autonomous agent, would endorse in this situation. However, if they had truly set themselves the goal to act otherwise than injecting themselves with the drug, on this specific occasion, then they could not also set themselves the contradictory goal of taking the drug, on this specific occasion. Therefore, they cannot intentionally take the drug, if they truly have endorsed the alternative to act otherwise. Buss (1994) suggests that the misunderstanding that they can intentionally do so is generally engendered by a confusion between attitudes towards an act-*type*, and attitudes towards *tokens* of that act-type. An individual may see many considerations against the act-type of injecting themselves with the drug. However, within the situation in which they find themselves, say a situation in which they are aware that their desire for the drug is so strong that it will eventually lead them to administer the drug to themselves against their will, the act-*token* of taking the drug might be their best response to the unfortunate circumstances. Alternatively, if they are led to take the drug contrary to the dictates of their goal-setting facility, then not only do they fail to act autonomously, but they fail to act intentionally at all.

Buss uses her arguments to make a case against 'endorsement'-type accounts of autonomy, which includes accounts like Frankfurt's (1971) higher-order desires account, and Watson's (2004) account based around well-considered values. However, the broad point can be applied to each of the various versions of the 'hierarchical' view: if the dominant aspect of the self under the hierarchical view, such as their decisive higher-order volitions, or their core values, or their rational judgement, represents the agent's own point of view on the situation at hand, then action that misaligns with this cannot be understood to exhibit the conclusion of the agent's own goals, and accordingly cannot be understood to be the agent's intentional action at all. Furthermore, as Buss (2013) suggests elsewhere, these considerations can be extended to action in general: if a reflexive agent rejects the power of those forces that lead them to behave in a certain way, then the resulting behaviour does not appear to be a manifestation of their agency. Therefore, the kind of alienation that contrasts with each of Frankfurt's explications of identification suggests that the resulting behaviour is not an instantiation of the agency of the relevant individual. Furthermore, I would add, if an individual's character, or set of core commitments, represents the outlook of the authoritative self, then when contradictory impulses successfully issue in behaviour, it is not clear that we should ascribe the outcome to the power of the agent, as their action.

In relation to the point around it being difficult to understand why the authoritative agent would *allow* such forces to overwhelm them, it seems that when the authoritative aspects of the self are confronted with motivational forces that run contrary to their dictates, there are two possible responses. One is that they could accommodate such forces, by incorporating them into their response to the situation. Perhaps their best judgement, or their best application of their values to the situation at hand, or their decisive higher-order volition about which desire they are to act upon, might adjust according to their understanding of this countervailing motivational force. It might be that, given the pain and the frustration of this impulse, it is best for them to adjust their response to the given situation at hand, according to the outlook of the authoritative aspect of themselves. It might even be, as suggested by Buss (1994) as one interpretation of the case of the unwilling addict, that they are aware that eventually this countervailing impulse will become irresistible, so it is best, even according to their authoritative outlook, simply to yield to it now.

Alternatively, they may remain firm in their resolution about the way to respond to the situation at hand. In this case, if the alien force does successfully lead them to behave in a certain way, it can only be by overwhelming the authoritative self. The authoritative aspect of

the self is hijacked by a force it is unable to resist. In this case, it is hardly convincing that the resulting behaviour is the agent's own. If the authoritative aspect of the self is a successful authority, then it will make the best it can of the motivational influences it finds itself with. If it does not, then, for a 'hierarchical' view, the only convincing explanation for this is that the self is overwhelmed. However, such behaviour does not seem to fulfil the minimal requirements to count as the agent's own action at all.

In defence of the 'hierarchical' views, it might be countered that the above arguments misconstrue hierarchical views in the role they assign to the 'core' aspects of the self that determine the conditions for autonomy. Whilst 'hierarchical' accounts do, it might be admitted, isolate certain aspects of the self that they deem to be more truly the agent's own, or more crucial when it comes to the identity of the self, they do not deny that other aspects form part of the self, too. Perhaps an intentional action might be the manifestation of some desire of the agent, and hence can be understood to issue from the agent as their action, yet it is in conflict with some more core and central element of the self – in this way intentional but nonautonomous action could be possible under a hierarchical account. Returning to Buss, the manifestation of an all-things-considered preference seems to be an especially strong condition to apply for intentional action, and we might wish instead to move to an alternative account, for instance Hyman's (2015, chapter 5) account of intentional action. For Hyman, an intentional action is the manifestation of a desire – this apparently provides a much weaker condition for the requirements of intentional action than Buss' notion of an all-things-considered preference.

However, the less unique and predominant the role assigned to the relevant 'core' aspect of the self, the less convincing that it is the sole benchmark for when an individual's behaviour is self-determined. As we recognise the validity of alternative aspects of the self as the possible underpinnings of an agent's action, it is harder to deny the claim that they also present relevant aspects of the agent's relation to a situation. The greater the status assigned to motivational influences that fall outside of the 'core' self, the less convincing it is that these influences do not also have a role to play when it comes to isolating self-determined action, or that these influences are not relevant aspects of the self when it comes to determining the conditions of autonomy.

Altogether, when it comes to 'hierarchical' accounts, the distinction they draw between autonomous action and action in general is ultimately untenable. By the defining the outlook

of the authoritative self in a limited and exclusive way, they undermine the idea that alternative possible motivational underpinnings of action can issue in behaviour that can be ascribed to the agent as a whole. This is problematic due to the convincing and intuitive examples of cases of nonautonomous action, including nonautonomous intentional action. It also means that they fail to show how autonomous action can form an interesting and illuminating category in its own right, aside from action in general, or intentional action in general.

Criticism (2): Integration

Given that a self can come apart in the way implied by 'hierarchical' accounts, in which the dictates of the authoritative aspects of the self conflict with other motivational forces, yet the latter are still effective in leading an agent to action, 'hierarchical' accounts also fail to explain the mechanism by which the authoritative self successfully self-governs, in the case in which the conditions of autonomous action are fulfilled. Unless all human action or behaviour is to be autonomous, 'hierarchical' accounts must grant that motivational forces that run contrary to the dictates of the authoritative aspects of the self also have the power to lead an agent to behave in a certain way. However, few theorists would deny that autonomous action is also possible, and generally autonomy is understood to express the default, and normal, condition of an agent. The authoritative aspects of the self are still the true authority, that represent the agent's own point of view on a situation, and generally the rest of the self is meant to align with them. In this case, however, 'hierarchical' theorists need to explain the mechanism by which the authority of those aspects of the self is normally successfully wielded. They need to explain how the authoritative self is able to gain traction over the self as a whole.

Buss (2013) alludes to such an issue, though in the context of the conditions of agency in general, as opposed to the conditions of autonomous agency specifically. As mentioned above, Buss suggests that if an agent is alienated from the motivational sources of their behaviour, a phenomenon alluded to by authors such as Frankfurt (1971, 2002), this undermines the view that the resulting behaviour can be a manifestation of their agency, or the agent's own action. We should therefore deny the view, furthered by theorists such as Wallace (2001), that rational beings succeed in acting as long as their behaviour is instrumentally rational and goal-directed. In contrast, Buss argues that an adequate account of

action must instead involve the person identifying with or endorsing the goals that her practical impulses lead her to pursue. A person does not act if they are overcome by psychic forces which they, as a rational being, ultimately reject. For Buss, insofar as we, as rational beings, take a stance on what *we are doing*, we endorse what we are doing, accepting the goals of our behaviour as our goals. Otherwise, the resulting behaviour does not qualify as our action – it is not our own doing. Therefore, an adequate account of our actions will have to account for the self-relation involved in our action, that means that we endorse the psychic origins of our behaviour.

On the other hand, for accounts of rational agency that do recognise the need for a rational agent to identify with her practical impulses, Buss takes issue with those that fail to adequately explain how the two aspects of rational agency are able to come together, such that an agent can act. For Buss, this occurs in accounts which assign entirely different roles to an agent's reason and her practical impulses in the production of action. For instance, under Korsgaard's (1996) authority-based conception of rational agency, the relation that holds between an agent's rational nature and her practical impulses is an authority relation – agency is secured when a rational being exerts authority over her motivational impulses, by bringing it about that the motivational impulses obey the laws that are constitutive of her rationality. However, for Buss, it is then mysterious how the laws of the agent's reason can be authoritatively binding on the motives that are independent of these laws. If the two sides of the self-relation have such distinct concerns, it is not clear why the nonrational side of the self-relation has any motivation to obey the rational side. Altogether, Buss sets out the following restriction on an adequate account of rational agency (p14):

'It must explain how

> (i)   a person's assumptions (conscious or unconscious) about what she has reason to do and
>
> (ii)   her nonrational impulses (whatever inclinations, desires, etc. are neither constitutive of her rationality nor the product of her reasoning)

can be sufficiently unified, or integrated, to constitute two aspects of a single point of view on her behaviour.'

For Buss, the capacity for self-alienation means that rational agency requires a form of integration within the agent, such that the concerns of the disparate parts of the soul have some common ground. Buss' account suggests a predominant role for the rational aspect of

the soul – it is the rational aspect of the soul that is able to take a stance in relation to the agent's motivational impulses, and determine whether they are to qualify as the agent's own, or not. In this sense, there is a background assumption of a 'hierarchical' type of structure. Moreover, her arguments apply to the conditions of agency in general, as opposed to the conditions of autonomous agency. However, her concerns can be extended to a critique of 'hierarchical' accounts of personal autonomy.

Such accounts presume that a core aspect of the self, such as the agent's set of higher-order volitions, or a set of core commitments that compose a character, or their best judgement about what to do in a situation, represents the outlook of the authoritative self, and that when the conditions for autonomy are fulfilled, this aspect of the self successfully wields authority and leads the individual to action. However, the mechanism by which it does so can also be bypassed by alien forces, such as, for instance, a first-order desire that the agent is alienated from. Otherwise, nonautonomous action would not be possible. The relevant mechanism is therefore not infallible, though it will operate successfully in the normal case. However, it is not then clear why the rest of the self, defined in a manner that separates it from the authoritative aspects of the self, should have any particular concern for the dictates of the authoritative self. It is a limited aspect of the self, such as an agent's rational nature, or a limited set of aspects of the self, such as their decisive higher-order volitions or their core commitments, that is deemed relevant for securing the conditions of autonomy. In maintaining this, but allowing that other motivational forces can operate independently of these aspects, 'hierarchical' accounts divide the self in a manner that does not effectively explain how that divided self can also come together.

It is not clear why a given first-order contrary desire, should be brought to align with an agent's decisive higher-order volition that runs counter to it. Moreover, in closer connection to Buss' observations, if an agent's rational nature is the authoritative aspect of the self, but that self can also be moved independently of that rational nature, for instance by desires that do not correspond to reasons, it is not clear why the independent desiderative aspect of the self should be brought to obey the rational aspect. Its concerns are distinct concerns. Finally, if the authoritative self is represented by an agent's set of core commitments or values, but the self can also be moved by contrary urges and impulses that contradict with them, it is not clear why the core commitments or values should have any special capacity to bring the rest of the self into line. Once again, the disparate aspects of the self seem responsive to different concerns and types of demands.

One problem, as alluded to by Buss, is how the divided soul is to come together in action – if the authoritative aspect of the soul has distinct concerns to the rest, then it is unclear how it is able to move the rest. On the other hand, if the rest operates independently to lead the individual to action, in such a manner that the authoritative aspect of the soul rejects, it is not clear that the resulting behaviour can be appropriately labelled as the agent's own action (see criticism (1) above – the possibility of nonautonomous action). On the other hand, even if we are able to solve the problem in relation to action, there is still the problem in relation to autonomous action: it is unclear why the dominant aspects of the self, under a hierarchical account, should have any special influence over the rest of the self, that is able to issue in action independently of them.

It may be countered that the authoritative aspect of the self has special authority simply in being at the top of the hierarchy, and the agent's final view on the situation at hand. For instance, for Frankfurt (1971), when an agent makes a decisive higher-order commitment, this 'resounds' up through the potentially endless hierarchy of further identifications. It is the agent's own point of view on the matter at hand, and in this sense, it might be countered, surely it has authority over the rest of the self as a whole. However, for hierarchical accounts this authority is simply left mysterious, and unexplained, given that they grant that the authoritative aspect of the self is not the only possible underpinning of action. Moreover, as I will touch upon in the next two chapters, it is not clear why this particular stance should have any special claim to represent the position of the authoritative self. Frankfurt (1971) might argue that it is to do with what makes us essentially persons, and theorists that identify the rational part of the soul as the top of the hierarchy (e.g. Raz 2002) may argue that we have an essentially rational nature. However, we are surely other things too, else conflicting behaviour that issues from other parts of ourselves – nonautonomous action, under hierarchical accounts – could not be appropriately labelled as our action at all. Moreover, as I will discuss in chapter four, the claim of any one part of the self to represent the agent's final say on a matter is ultimately unstable. Therefore, limiting the outlook of the authoritative self to one distinct element in such a way underestimates the complexity of human nature.

For Buss, the solution to the problem of integration when it comes to rational agency is that both the rational part of the soul and the nonrational part of the soul regard themselves as responding to mind-independent reasons to behave in a certain way. However, this then assumes that the agent takes themselves to be responding to an objective normative reality that is independent of the agent, whenever they constitute themselves in action, which

presents quite a strong constraint on the way in which the agent must interpret the decision-situation they face. Since autonomy as self-governance is a *self*-relation, that ideally should be consistent with a wide range of interpretations of the good, this is not an adequate solution for the problem as applied to autonomous agency – we must accept that the different parts of the soul can be responding to different matters and concerns.

Altogether, whilst Buss' solution is not fully satisfactory, and Buss is targeting the basic requirements of agency, as opposed to the basic requirements of autonomous agency, she still points towards a problem that needs to be addressed by an adequate account of autonomous action. Such an account must explain how the governing aspect of the self – for instance the agent's higher-order desires, her acceptances and authorised preferences, the aspects of the self that compose her character or narrative unity, or her all-things-considered judgement about the good to pursue in a situation – is able to wield authority over the governed aspect of the self. It must explain how the part of the self which can issue in action without the endorsement, approval, or sanction of the agent as a whole (for example, an agent's set of first-order desires), can be brought to obey the authoritative aspect of the self (for example, an agent's second-order volitions). This is not a challenge that 'hierarchical' accounts are adequately equipped to meet.

Alternatively, it might be admitted that the relevant aspects of the self are so distinct that the dominant aspects of the self have no influence over the rest. In this case, however, then, the rest seems to be no more than a set of additional external impediments for the agent to take into account in deciding how to act. For instance, say we hold a reasons-responsiveness conception of autonomous action, but the agent has a desiderative faculty that is completely immune to reason. In this case, it seems as though their desiderative faculty can only pull them in one direction or another in a manner that entirely bypasses the authoritative aspects of the self, when it comes to autonomy. This links to the other problem that accounts of personal autonomy need to address: explaining the possibility of nonautonomous action or agency. In such cases, it is not very convincing that the resulting behaviour is truly the agent's own.

Before moving on, it is worth touching upon a possible further extension of Buss' remarks about reflexivity. Buss argues that human reflexivity – the ability to reflect upon and assess our own mental states - implies an interaction between two different elements involved in acting for a reason (p14):

<ol type="i">
a person's assumptions (conscious or unconscious) about what she has reason to do and
her nonrational impulses (whatever inclinations, desires, etc. are neither constitutive of her rationality nor the product of her reasoning)

The interaction is that we can distance ourselves from our practical impulses and take a stance in relation to them, presumably with reference to reasons. (This is what grants us the capacity for self-alienation). However, I contend that whichever benchmark we are using to assess our motives and inclinations by – say, a set of reasons, or normative criteria – is itself something which we could question and take a stance in relation to, by reference to other standards and benchmarks. We may be able to assess our motives and inclinations on the basis of reasons, but these reasons themselves are the potential subject of assessment. This relates to the regress concerns around Frankfurt's (1971) higher-order desires account of autonomy alluded to above, as pointed out by authors such as Watson (2004): for Frankfurt, we are autonomous when we act on the basis of our higher-order volitions about how to behave – our desires about which desires are to be effective in moving us to action. An agent identifies with a desire that is the object of a higher-order volition. However, if this is the standard for identification or internality with respect to desires, then it seems it needs to be applied at the second-order stage of desires, too. It is not convincing that an aspect of the self that the agent does not already identify with can be used to ground internality with respect to mental states.

Altogether, whichever aspect is chosen as the 'authoritative' aspect of the self, it seems that the agent can have the capacity to distance themselves from it, and take a stance with respect to it. This suggests that the regress issues around Frankfurt's higher-order desires account of autonomous action can generalise to other versions of the 'hierarchical' account too. All suggest a final stance that represents the agent's own, or true, point of view. However, whichever way this stance is defined, it is always merely a further outlook, that the conditions for personal autonomy can always be applied back to. For instance, the reasons that inform the stance of the agent in a reasons-responsiveness account can always be assessed by reference to further reasons. This suggests an instability within the structure of 'hierarchical' accounts of the self-relation involved in personal autonomy. I will discuss related concerns to these in chapter four, in which I put forward a positive argument in favour of the kind of self-relation implied by a horizontal, as opposed to hierarchical, account of personal autonomy.

<u>Conclusion</u>

Altogether, 'hierarchical' accounts of personal autonomy offer an intuitive and simple model of autonomous action, by delineating certain limited aspects of the self that represent the 'true' self or the core self, and defining autonomous action as action that aligns with them. However, these hierarchical accounts fall prey to two shortfalls that I have detailed in this chapter: they fail to explain how nonautonomous action, including nonautonomous intentional action, can be possible, and they fail to illuminate the mechanism by which the core aspect(s) of the self are able to wield authority over the rest, such that the rest is responsive to them and the two (or more) sides of the soul are integrated. In the next chapter, I will outline the shape of a 'horizontal' alternative, as a structure for an account of personal autonomy. Following that, in chapter four I will detail the ways in which a 'horizontal' account can bypass the challenges faced by 'hierarchical' accounts that I have discussed in this chapter, as well as presenting a positive argument in favour of it.

Chapter 3: Outline of the Horizontal Account

Introduction

As discussed in the previous chapter, traditional 'hierarchical' accounts of personal autonomy fail to convincingly allow for the possibility of nonautonomous action, or to illuminate the mechanism by which the conditions of autonomous action can be fulfilled.

In this chapter, I will outline the contours of a 'horizontal' alternative. The horizontal account does not delineate one limited aspect of the self that exerts authority, when an individual is autonomous. Rather, the self as a whole, together with its relation to the outside world, determines the conditions of autonomy. First, I will discuss the way in which the horizontal account intends to 'level' the hierarchy implied by alternative accounts. In doing so, I will consider dialogical models of the self, and the way in which they might be adopted by a horizontal account of autonomy (Hermans 2001). Next, I will discuss some ways in which the conditions for autonomy might be determined, under a horizontal model. I will consider some of the strengths and weaknesses of these various approaches.

Whereas this chapter focuses on the shape of a horizontal account, in the next chapter I will run through a positive argument in favour of it: the self-as-deliberator, or self-as-processor, inevitably undergoes its processes on the basis of some outlook, or some materials. Therefore, whenever we deliberate or reach a decision, we do so on the basis of an outlook that can be itself subjected to scrutiny. In the next chapter I will also describe the ways in which the horizontal account of autonomy is able to bypass the shortfalls of 'hierarchical' accounts that I discussed in chapter 1.

Levelling the Hierarchy

As discussed in the previous chapter, different 'hierarchical' accounts limit the purview of the authoritative self in different ways. Frankfurt's (1971) account, for instance, limits the scope of the authoritative self to an individual's set of decisive higher-order commitments – their decisive volitions that specific first-order desires be the ones to move them to action (1971, p16). The individual is autonomous when their first-order effective desires (the desires that move, or will or would move, them to action) are in line with their decisive higher-order

26

volitions about those desires (their higher-order desires that certain desires be the ones to lead them to action). Therefore, the aspect of the self that wields control when an agent is autonomous is this set of higher-order volitions. Contradictory desires fall out of the purview of the authoritative self.

Alternative accounts emphasise the individual's capacity to respond to reasons, or their capacity to respond to reasons as they understand them. For instance, Raz (2002) emphasises an individual's responsiveness to their understanding of the reasons available. For such accounts, the authoritative self is the rational self – the self that recognises and responds to reasons, or else reasons as they understand them. (For certain authors, this aspect of the self is responsive to mind-independent reasons – see Buss 1994). Motivational impulses that contradict with the outcome of rational assessment fall outside the authoritative self. In contrast, character accounts emphasise core desires, commitments, or values that are said to compose an individual's character (see e.g. Dworkin 1970) – for such accounts, desires, commitments, values or beliefs not contained within this subset are not part of the authoritative self.

In contrast, a 'horizontal' account of personal autonomy denies that the 'authoritative' aspect of the self can be contained to one distinct segment in this way, such as the agent's set of decisive higher-order volitions, or their rational nature, or core commitments that compose their 'character'. Instead, all of these aspects and more can have a role to play when it comes to securing autonomy. The 'true self' is not one core aspect of the self, but rather is the self in its entirety. This includes all of an individual's desires, beliefs, and values, including desires about desires and beliefs about reasons. It includes all aspects of the individual's outlook about the world, together with all aspects of the individual's view upon themselves, including their view about these views and outlooks (if relevant) (i.e. mental states of a higher-order). It also includes their capacities and abilities, such as their capacity to recognise and respond to reasons, and their ability to process new information. In this sense, the 'hierarchy' is levelled, because all the aspects of the self formerly picked out by 'hierarchical' theorists are absolved of their unique status. Instead, the state of the self in its entirety becomes relevant for determining the conditions of autonomy. This is the self as rational subject, as first-order desirer, as second-order desirer, along with a host of other manifestations. For a specific individual, this might include the self as mother, as teacher, as lover of chess, as rational agent, and as reflexive agent.

One way of illuminating the 'horizontal', levelled out self, is through a dialogic conception of the self. Dialogic conceptions of the self emphasise the disunity of the self, and the ways in which different aspects of the self can interact in a manner analogous to conversation. However, unlike 'hierarchical' accounts of autonomy, dialogic models do not isolate one part as the dominant part, or the 'true self'.

Dialogic conceptions suggest that the self can be broken into parts, and that the dictates of these parts may conflict with one another. This feature of dialogic conceptions of the self has parallels with views of personal autonomy that involve isolating distinct aspects of the self, that wield control when an agent is autonomous. Such views of autonomy suggest that the self can be broken up, in a way that opens the possibility of conflict. Take Frankfurt's unwilling addict (1971). He has a first-order desire to take the drug (as a result of his addiction) together with a first-order desire not to take the drug (in recognition of a potentially large number of considerations against taking the drug). He also has a second-order volition that his desire not to take the drug is the one that leads him to action (deeming the considerations against taking the drug to be superior to the satisfaction involved in taking it). In this case, then, his second-order volition, the part of the self relevant for determining the conditions of autonomy (for Frankfurt (1971)), is in conflict with a first-order desire to take the drug. The individual will be autonomous if this second-order volition wins out over the first-order desire to take the drug, and his first-order desire not to take the drug is the one that is effective in determining how he behaves.

It is possible to conceive of a self similarly divided into parts, some of which may conflict with one another, but without the assumption that one specific element is authoritative. The dialogical self, as discussed by Herman (2001), provides a useful framework for conceptualising such a structure. Inspired by work from James (1890) and Bakhtin (1973), Herman discusses Hermans, Kempen and Van Loon's (1992) conception of the self based around, as Herman puts it, 'a dynamic multiplicity of relatively autonomous I-positions' (p248). The 'I' is loosely based on James' conception of the 'self-as-knower' (p246). The self-as-knower is a continuous, distinct, and volitional being. It is continuous in that it maintains a sense of personal identity over time, distinct in that it is subjective, which renders it a sense of individuality, and volitional in that it is able to process experience in an active way. This contrasts with the Me or self-as-known, the empirical self, which contains the empirical elements that are understood to belong to oneself.

For James, the 'I' might appropriate distinct aspects of the Me at different times, for instance the self as 'philosopher' or the self as 'philanthropist' (James 1890 p309-310). Similarly, for Hermans, Kempen and Van Loon, the 'I' can fluctuate between different positions within the self, positions which may conflict with one another (p248). Moreover, inspired by Bakhtin's (1973) work on literary theory and polyphony in novels, in which multiple opposing views of the world are contained within the world of a single individual, Hermans et al.'s I is able to imaginatively assign each position a voice, enabling dialogical or conversational relationships to form between the different positions taken up by the I. There is a continuity of experience between the I in each of the relevant positions, because, whichever position it takes, it is still one and the same self. However, there is also a discontinuity between the I in each of its separate positions – each represents a different voice with a different set of demands, and they may well be opposed to one another. The I in one position is able to question and criticise the I in another position.

Like hierarchical conceptions of autonomy, such a dialogical conception of the self maintains the sense that the I can be broken into parts. However, in contrast, it does not identify one unique part, or a selection of special parts, which are to be delineated as the 'true' self. Rather, each aspect of Me can stake its claim as a distinct, autonomous I, can demand to be heard, and the various aspects of Me form different positions which can be taken up by the single I at different points in time (or even at the same time, when the distinct I-positions are in conflict). It is not that there is one, true, permanent position for the 'I' to hold, which, under a hierarchical account of personal autonomy, determines whether autonomy is secured. Neither is it the case that there is a limited set of such true, permanent (or semi-permanent) positions, each with a similar structure, for instance an agent's set of higher-order volitions (Frankfurt 1971), or the agent as responder to mind-independent reasons (Buss 1994). Rather, the self can temporarily take on different perspectives, which may be opposed to one another.

Applying to Frankfurt's unwilling addict, it might be that the 'I' can take on the position of the desire to take the drug, informed by considerations such as the prospect of short-term pleasure and the cessation of pain of withdrawal symptoms. However, it can also take on the position of the desire not to take the drug, informed by contrary considerations such as desires to fulfil family and work commitments. These contrary stances can compete and conflict with one another. However, both stances are still internal to the self, as broadly defined under the horizontal account. I would like to draw from the dialogical conception of the self the understanding that the self can contain different competing positions, without the

notion that a limited subset are to be labelled as internal and the rest external. Instead, under the horizontal conception of autonomy, all such positions are different aspects of the authoritative self. Not all elements of the self will be contained within any one perspective that the self takes on. However, it is not to say that the rest fails to form part of the 'true self'. It is possible for the self to take on different perspectives with outlooks informed by different elements of the self, and higher perspectives that adjudicate between and incorporate the materials of multiple different lower perspectives.

As well as aligning itself with first-order mental states, notably first-order desires, that have a worldly object, the self is also able to align itself with second-order mental states. In doing so, it adopts a position in relation to its own stances to the world. The object of these second-order mental states is an attitude of the agent, such as a belief or desire. For the unwilling addict, the self aligns itself with the second-order volition that their first-order desire not to take the drug be the one that is effective in leading them to action. In doing so, the self applies its capacity for reflexivity. However, according to the dialogical conception of the self as applied here, it is not to say that in doing so, it reaches a final, ultimately decisive position, that sets in place the ultimate nature of the true self. As long as the contrary first-order desires are still there, it is possible for the self to identify with their position, and no position has the claim to be the one, 'true', ultimately authoritative self. In this sense, we level the hierarchy. Neither one particular aspect of the self, nor, in particular, the self at the second or further stage of being a reflexive being, are granted the status of the 'true' self, alignment with which secures an individual agent's autonomy. Even when the contents of multiple perspectives are incorporated into a higher perspective, e.g. a higher-order perspective that absorbs and adjudicates between the perspectives both of the desire to take the drug and the desire not to take the drug, this is never the final or complete outlook of the authoritative self – there will be further elements of the self that are not included but still form part of the entirety of the authoritative, holistic self.

Each 'I'-position involves a perspective that may be rich in content but is limited. It may also be critiqued and assessed from the position of other perspectives contained within the same self. Whilst adopting a perspective, the contents of that perspective are *applied*, rather than brought into scrutiny. However, from another perspective, the contents of that former perspective can be assessed, in relation to other outlooks and other sets of standards.

This assessment from another perspective may involve assessing the worldly object of scrutiny contained within the first perspective, such as the object of a desire, or a possible course of action, by reference to alternative desires, beliefs, and reasons. Alternatively, it may involve taking a stance in relation to the aspects of the self that are employed in the former perspective – it may involve assessing the desire qua desire (as opposed to assessing the object of the desire), or assessing the intention to carry out a particular course of action. In this latter case, the deliberation or reflection takes place at the second order level – the object of reflection or deliberation is not a representational object in the world, but rather an aspect of the self, potentially with representational content (such as a belief or an intention). However, in adopting a perspective of such a second order kind, the self will still be applying an outlook informed by various beliefs, desires, and views about reasons, different to those contained in the former perspective. Even when reflection takes place at the second-order or further-order level, it, whether consciously or unconsciously, involves the application of materials which can themselves be subjected to scrutiny, from other perspectives.

This is not to say that the human capacity for reflexivity has no role to play when it comes to the autonomous individual. However, even when the capacity for reflexivity is applied, and an individual takes a stance in relation to themselves as opposed to a stance towards an aspect of the world, there must be some materials by which the relevant scrutiny is carried out. Furthermore, even if the self takes a position between multiple perspectives, weighing up the demands and views expressed by a number of different positions, it will still be doing so from a further perspective, which will be inevitably limited. This latter point has the propensity to be overlooked in discussions of the notion of an 'all-things-considered' judgement, which for some authors underlies the possibility of autonomy.

For instance, as discussed in the previous chapter, there is Denis McManus' (2019) interpretation of Heidegger's (SZ) notion of authenticity or 'ownness' according to which authentic action is the manifestation of 'all-things-considered judgements' (p1194). As discussed previously, for McManus, such a judgement involves an individual considering the entire situation as it presents itself to them, including all its normative pulls in one direction or another, and accordingly deciding how to act. This contrasts with action that reflects a mere 'Q judgement' (p1194) – a judgement carried out from the point of view of some particular goal, or some particular role. This has parallels with Buss' (1994) view that intentional action is the manifestation of all-things-considered judgements about how to behave, as discussed in the previous chapter. However, in this case ATC judgement is the

marker of a particular species of action – authentic action – as opposed to intentional action in general.

However, the individual carrying out such an 'ATC' judgement must inevitably be doing so from some perspective. From this perspective, the individual will be applying principles, standards, reasons, beliefs, or desires, that will not be the object of scrutiny, but rather the materials by which that scrutiny is carried out. (See the next chapter for further consideration in favour of this). The materials by which the scrutiny is carried out can always themselves be subjected to scrutiny. Moreover, as a result of human limitations and practical barriers, the view adopted in the application of an 'ATC' judgement can never truly be complete. I therefore contend that as for a higher-order attitude, neither can such an 'ATC' judgement offer itself as the top of the hierarchy, or the manifestation of the single, true, authentic self that determines the conditions for autonomous action. Similarly, each of the other versions of the 'hierarchical' account discussed in the previous chapter oversimplifies the picture when it comes to the relevant authoritative aspects of the self for autonomy. On the contrary, I argue that each possible motivational underpinning of action, or belief whether about reasons or otherwise, is an aspect of the authoritative, holistic self, and has a possible role to play when it comes to securing the conditions of autonomy. With support from the critique offered in the previous chapter, I therefore argue that we should be moving towards a horizontal account of the self-relation involved in autonomy, where the self as a whole is relevant for determining the conditions of autonomy. In the next section, I will outline a number of ways in which the detail of the horizontal account could be filled out.

Horizontal Accounts of Personal Autonomy

'All-things-considered' judgements and human reflexivity reconsidered

As discussed, an 'all-things-considered' judgement can never truly be all-things-considered. It is false to argue that a self can ever fully account for what a situation calls for in its entirety. When a self makes a judgement, whether that is in relation to an object in the world (e.g. in adopting a first-order desire, for a specific outcome), or in relation to an aspect of the self (e.g. in adopting a second-order desire, about a desire), it will be incorporating a limited range of internal information, and a limited range of external information. When I say 'internal information', I am referring to the agent's own beliefs, desires, intentions, values, emotions, and so on. When I say 'external information', I am referring to facts in the world.

When a self makes a judgement, it will inevitably be ignorant of relevant facts. Moreover, neither will it be incorporating all the information that is on some level aware of. This is the information contained in the self's own attitudes that are pertinent to the judgement, the stances that the self has taken to various aspects of the world (whether in the form of beliefs, desires, values etc.). This will not all be fully incorporated into the judgement. In reality, we may be able to move beyond McManus' (2019) mere simple 'Q judgements' (p1194), to wider judgements that incorporate a greater range of material. However, it is unrealistic that we will ever get to a true 'all-things-considered judgement', as opposed to successively more sophisticated Q judgements.

However, McManus (2019) (and Heidegger SZ, under McManus' interpretation), touch upon a compelling intuition: that an individual is more autonomous when a greater range of relevant factors are brought to bear upon a given judgement. An 'all-things-considered' judgement is meant to be a judgement in which the self contemplates what the situation calls for in its entirety. This may be an unrealistic picture, but it is more convincing that the self is autonomous when it takes into account a greater number of its desires, and incorporates a wider view in relation to a situation. This is in contrast to actions and decisions in which the self acts on some passing impulse, whilst ignoring their wider set of values and a potentially large number of reasons that run counter to their carrying out that thing. When an agent takes on a wider view, it is more convincing that their action represents themselves and can be labelled their own.

For instance, say an addict who truly cares for their family and their work above all else, and is aware on some level that they cannot succumb to the drug without falling into a pattern of behaviour that will be greatly detrimental to these aspects of their life, nonetheless acts on their desire to take the drug whilst ignoring or overlooking all these other considerations. In this case, intuitively, they are not autonomous. A horizontal account can explain this by reference to the action's relation to the self as a whole, and potentially moreover to the world, as interpreted by that self. The action only reflects a very limited aspect of the self, widely understood under the horizontal account as incorporating its beliefs, including beliefs about reasons, desires, values, and more. Moreover, the action fails to incorporate a large number of considerations that the self would take to count against behaving in this way, if considered, whether these are considerations that the self is on some level aware of (e.g. latent beliefs), or whether these are factors that the agent is truly ignorant of. Under this picture, an autonomy-undermining influence is one that leads an individual's perspective to become curtailed in

such a way, for instance the power of an addictive substance and the strength of the individual's desire for the drug.

When it comes to reflexivity, intuitively it is also more convincing that the self is autonomous when its higher-order attitudes are incorporated into the decision about how to behave. However, it is worth considering why this might be the case. It is worth noting that when a higher-order attitude is incorporated into the decision-making, not only are a wider subsection of the attitudes of the self, as a naturally reflexive agent, brought to bear upon the judgement, but also a wider range of information about the world. Higher-order attitudes are formed when the self takes a somewhat detached perspective upon itself, as an additional object in the world that it can make predictions about and wish to be swayed in one direction upon another. When a self forms a higher-order attitude, it incorporates its own weaknesses (for instance, inabilities to follow through on intentions), as relevant factors in the world to be incorporated in deciding what to do, or deciding what it would like to be (in the case of taking stances towards its own mental attitudes or states). Moreover, reflexivity tends to involve the agent 'stepping back' from their immediate inclinations, and assessing them against other values that they hold, or long-term plans, or reasons for and against the immediate gratification of their desires. In this way, the human capacity for reflexivity opens the potential for a greater range of relevant factors to be brought to bear upon a decision or judgement, widening the perspective adopted by the agent.

Example 1: Heyman: Addiction as Choice

Specific examples of autonomy-deprivation suggest similar conclusions about the structure of high-autonomy states versus low-autonomy states. Take Heyman's (2010) account of addiction as choice. Heyman (2010, chapter 6) suggests that addiction involves myopic choice, in which individuals apply 'local bookkeeping' as opposed to 'global bookkeeping' when deciding what to do. Local bookkeeping involves choosing the action with the best returns at the moment of decision-making, as opposed to choosing to act in ways that will overall bring better returns over a longer period of time. The outcomes of local bookkeeping and global bookkeeping are likely to supply different answers about the best way to behave in a situation. Returning to the addiction example, applying local bookkeeping, a decision to take the drug now or not to take the drug now will incorporate the impact of taking the drug

as an isolated act. The benefits of taking the drug include an intense amount of short-term pleasure, or at least cessation of pain of withdrawal symptoms, in comparison with a lower amount of immediate gratification that can be gained from alternative nondrug activities, such as attending to various life commitments, relationships, and hobbies. Over time, as the addict neglects the other aspects of their lives that the nondrug activities relate to, these nondrug activities also diminish in value, for instance they may lose a relationship. The alternatives to drug-taking become less appealing. Even if the pleasure from the drug diminishes over time, then, as they become dependent, taking the drug will continue to be the preferred option from the local bookkeeping perspective: in terms of immediate benefits versus disbenefits, it is better for them to choose to take the drug now.

Global bookkeeping, on the other hand, will assess the impact of taking the drug as an act-type in similar situations over a long period of time. Over time, the marginal negative detrimental impacts on the individual's long-term goals from each instance of drug-taking will compound in such a way as to undermine the aspects of the individual's life that they most care about. The pleasure from the drug will be diminishing over time, and the value of these multiple but diminishing short-term pleasures does not offset the cost of the individual falling short on their long-term goals and commitments. The global equilibrium, therefore, which considers the benefits of taking the drug versus not taking the drug in terms of a bundle of choices over time, suggests that it is better to abstain from the drug.

It is striking that the adoption of one perspective rather than the other – the local perspective versus the global perspective – brings about entirely contradictory results in terms of the correct choice for the addict to make. The local bookkeeping perspective brings the result that on each day, in say a 30 day bundle, the addict chooses to take the drug, whereas the global bookkeeping perspective brings about the result that the addict should abstain on all of the days (p125). Heyman uses his results to support the contention that addiction behaviour can be understood as voluntary – if a local perspective is adopted, addiction behaviour can be a intelligible reaction to the (immediate) expected consequences. However, the overall gains to be made in the global equilibrium are much greater than in the local equilibrium, suggesting that such an attenuated perspective produces long-term irrational results over time.

In support of his view of addiction, Heyman (chapter 3) points to empirical evidence which suggests that life events that lead an individual to broaden their perspective in such a way as to apply global bookkeeping, as opposed to local bookkeeping, can ultimately help an

individual break out of the addiction cycle. Some of his evidence is anecdotal: for instance, Patty was a cocaine addict as well as a single mother of two daughters (Waldorf et al., 1991). She used cocaine heavily for fifteen years, and sold drugs on the side to help fund her cocaine habit, alongside the cost of maintaining her family. However, when she was almost arrested one day, she decided to quit selling drugs. Without the money from dealing, she could no longer fund both her drug habit and food for her family, so she in turn gave up taking cocaine. A quote in which she outlines her rationale for this turn of events suggests that she considered her role of P.T.A. president, and how bad it would look if she got caught, together with the possibility of embarrassing her daughters, who were getting older and starting to have friends over. Without the money from dealing, it would be a case of feeding her daughters versus taking cocaine. The kinds of factors she references as leading to her quitting cocaine suggest that her breaking out of the addiction cycle was fuelled by events that led her to adopt a more global perspective in relation to her drug choices, as opposed to a local perspective. She considered her roles as both P.T.A. president and mother, and, as Heyman points out (chapter 6), social roles like these refer to patterns of behaviour, as opposed to individual choices – being a good mother, or a good P.T.A president, is not dependent on one action alone, but a series of actions over time. Such considerations suggest the adoption of global bookkeeping, as opposed to local bookkeeping.

Heyman also refers to studies concerning the success rates for various rehabilitation programmes. One programme, developed by Steve Higgins at the University of Vermont (p106, Higgins et al. 1991) involved two elements: counselling sessions that would help the patients develop their drug-free lifestyle, and vouchers that would be offered upon evidence of drug abstinence, following tests. The vouchers could be traded for goods like sports equipment and theatre tickets, and would increase in value as the period of abstinence increased in length. The results for one of the cocaine experiments suggested that around 70 percent of the individuals who received vouchers managed to remain drug-free for the first five weeks, compared to just 20 percent of a control group who just received traditional psychological counselling. This suggests that addicts were aided by a voucher scheme that helped to improve the perceived immediate benefits from abstaining from drugs, making abstinence a more appealing option within the local bookkeeping perspective. A later study (Higgins et all., 2000) investigated whether relapse would occur, once the voucher incentive was removed. Similar control and treatment groups were put in place as before. For each follow-up date, individuals who had been part of the voucher programme were more likely to

stay off drugs, even though they were no longer receiving vouchers. As Heyman puts it: 'healthy alternative activities can set in motion an upward spiral of increasingly positive consequences and increasing options' (p107). It seems that the vouchers supported the pursuit of activities that helped promote a more global perspective – the pursuit of goods with long-term benefits to be achieved over time. Altogether, the evidence Heyman gathers suggests that addiction, and the accompanying underpinning of autonomy, can be fuelled by a myopic, local bookkeeping approach to decision making.

Local bookkeeping involves incorporating a limited range of information compared to global bookkeeping, which brings a greater range of goods to bear when it comes to the action decision, including the ability to keep a job, and the ability to fulfil family and social commitments. Local bookkeeping may, admittedly, be rational in some sense. It correctly predicts that, on this particular occasion, the individual likely has more to gain from taking the drug. Moreover, local bookkeeping involves fewer calculative costs, and is timesaving. However, when it comes to addictive substances, and many other kinds of consumption decisions (chapter 6), the attenuated perspective involved in local bookkeeping brings about detrimental outcomes and is intuitively autonomy-undermining, for instance in the promotion of addiction behaviours. Global bookkeeping, on the other hand, contributes to a higher autonomy state, through the individual taking on a wider view and envisaging more possibilities: long-term possibilities that can be achieved only with the aid of repeated acts of restraint, that come with a high short-term cost. In the high-autonomy state, the individual factors a greater range of goods into their decision, including goods that can only be realised over long periods of time. The decision-scenario, as presented to them, involves more options, and a richer set of options, and includes a greater range of relevant factors to be incorporated into their decision-making.

As Heyman points out (chapter 6), it is possible for humans to spontaneously employ a global perspective, as well as for them to be led towards the adoption of a global perspective, and it seems that humans are the only species that are able to do this. Heyman suggests that this relates to the fact that the items involved in global choice are abstractions, such as the fulfilment of a role, or the achievement of a good that can only be materially obtained far into the future, through concerted repeated efforts. On the other hand, local choice generally concerns items that we can immediately perceive. The importance of the ability to take on the kind of rich perspective implied by a global perspective therefore helps explain why individuals with limited rational and cognitive abilities may be considered less autonomous.

It also links back to the perceived importance of human reflexive abilities and 'all-things-considered' judgements when it comes to autonomy. Human reflexive abilities imply a special additional capacity to reflect upon and assess items that do not form part of our immediate perceptual experience – our beliefs and desires, as opposed to the objects of those beliefs and desires. Moreover, judgements that involve a global perspective involve relevant additional abstract elements that might help a perspective approach the ideal of an 'all-things-considered' judgement. Altogether, it seems that local bookkeeping perspectives can be autonomy-undermining, as in the case in which they support addiction behaviour, whereas the richer and broader perspectives involved in global bookkeeping are contributory to higher levels of autonomy. Heyman's empirical evidence on addiction supports this picture.

Example 2: Jonathan Lear: Radical Hope

An alternative type of example of autonomy-deprivation is that described in Jonathan Lear's *Radical Hope* (2006, chapter 1), in which a group of people – the Crow Nation of north America – are deprived of autonomy when the normative concepts that were formerly used to direct their lives are rendered meaningless. This is via outside influence that destroys the conditions of their application. Once a warrior and hunting nation, their perspective and decision when it came to how to direct their lives was built around values and norms solidly rooted in success in those pursuits. However, once they were forced to move to a reservation in order to survive, they were no longer able to apply those normative concepts, since the material conditions of their application were destroyed.

For instance, their conception of courage was built around bravery on the battlefield, and defending their hunting land against outside military threats, such as those of the rival Sioux nation, who would threaten their claim to their hunting lands. A courageous warrior would defend the Crow territory to the death, planting a 'coup-stick' to indicate a point beyond which no combatant could go during battle, on pain of the Crow warrior defending that land right to the point of death (p34). Warriors who 'counted coups' in such a way – planting multiple coup-sticks – would be celebrated among the Crow people. However, given the evolution of historical events, eventually the Crow people realised (or surmised) that, given the threat of the encroaching European settlers, and the diminishing supply of buffalo to hunt, the new best way for them to protect themselves against existential threats was to make a deal with the European settlers.

The deal meant that they would get to keep (some of) their land, but would need to live on a reservation and embark on a new kind of life, deprived of their former nomadic, warrior-focused existence. Given this, concepts such as that of 'counting coups' and 'coup-sticks' were deprived of the conditions under which they were formerly meaningful. Their old conception of courage no longer made sense in the new world they found themselves in. More generally, the former normative framework around which they had built their lives, granting them a telos, was deprived of its former meaning, and in that way undermined. As Plenty Coups, a leader of the tribe, stated, in relation to the period after the buffalo went away: 'After this nothing happened.' (p2). This indicates the way in which cultural devastation deprived the Crow people of the normative framework that had rendered meaning to their lives. Following this period, they were disorientated and felt lost and confused, and could perceive little value in their activities.

Lear (2006, chapter 2) goes on to describe the way in which the Crow people were able to emerge from this crisis to an extent, by finding new meanings and new interpretations by which to understand their lives. However, the experience of the Crow people indicates another way in which an individual's (or a culture's) perspective can be autonomy-undermining: if the individual does not have access to frameworks of meaning around which to build their lives, their perspective, including of the options available to them, will once again be attenuated. 'Thick' concepts such as a filled-out conception of courage broaden an individual's scope of possibilities, that can be incorporated into decision-making. A concept like 'counting coups' is simply not applicable anymore, once the material and societal conditions of its application are no longer in place.

In contrast, as part of our membership in groups and societies, we are granted the opportunity to carry out meaningful actions that would not be meaningful if it were not for our membership in those groups and our ability to grasp the associated concepts. For instance, placing a piece of marked paper in a certain box only counts as the action of 'voting' under certain concrete conditions. Together with the material conditions of their application, the availability of normative concepts like these is determined by the collective consciousness of a society, and intersubjective agreement about what certain actions will mean. Without such possibilities, an individual simply has more limited scope in making a particular judgement. Once again, this is autonomy-undermining. To achieve a certain level of autonomy, we require a minimum set of options, and a minimum set of goods that can be factored into our

decision-making, and this is to a certain degree determined by the health and breadth of a society's 'thick' concepts.

Lessons for a Horizontal Account of Autonomy

The examples of autonomy-deprivation presented by Heyman and Lear suggest two ways in which autonomy can be truncated, and accordingly provide lessons about what a contrasting high-autonomy state might look like. Take Heyman's example of addiction involving myopic choice. The contrasting, high-autonomy state is one in which an individual makes a decision in which they account for the consequences of their action over the course of a lifetime, or at least over a significant period of time, as opposed to merely factoring in its immediate merits and dismerits (global bookkeeping as opposed to local bookkeeping). This involves incorporating a broader set of considerations in relation to their action, and taking a wider view in relation to a situation. As opposed to just conceptualising the possibilities available to them in the immediate future, they conceptualise a richer picture of long-term gains and possibilities. Moreover, by taking a wider view on the paths afforded to them by various immediate action options, and considering not just a single action but that action contextualised within a plan over the course of a lifetime, new possibilities are opened to them, possibilities that they can reasonably expect to achieve by the way in which they behave. To develop this richer conceptualisation of alternatives, they incorporate more relevant information than that contained in the short-term, local bookkeeping perspective, whether that is information that they were on some level aware of despite applying the local perspective, or whether that is entirely new information that they acquired as they began to apply the global perspective (as they broke out of the addiction cycle). This additional information is partly characterised by its more abstract nature, compared to information contained in the short-term, local perspective, including the possibility of fulfilling various obligations and roles, and the achievement of distant goods and goals.

In Lear's example, the low-autonomy state is one in which a people are deprived of the 'thick' normative concepts around which their lives and actions can be rendered meaningful. These concepts having been deprived of the material conditions for their application, the Crow people literally had fewer actions to choose from. For instance, 'counting coups' was no longer an intelligible act. With various actions deprived of their former meaning, even if the individual went through the motions of such former action types, they would not be

performing the same action as before, and they would not be able to achieve the same things by behaving in that way, as before. This reduced the scope of possibilities available to them, as they considered what to do. Without a rich normative framework around which to build their lives, there was little they could do that had relevance within a wider context. In this case, the corresponding high-autonomy state is one in which a culture has a healthy selection of 'thick' normative concepts, which provide options for the ways in which people can live their lives, and offer new possibilities for action (an action defined here not merely as a bodily movement, but rather in terms of its significance within a larger picture, which may include its consequences). This requires an individual having a healthy range of thick concepts available to them – they need to find themselves within a thriving culture, and also have the opportunity and understanding to ingratiate themselves within it. Once again, this means a more expansive range of materials is incorporated into the judgement about what to do.

In line with the observations around higher-order mental states, and 'all-things-considered' judgements, above, the kinds of high autonomy states that run contrary to the Heyman and Lear examples are ones in which a richer conceptualisation of the world, including the future possibilities available to one, is brought to bear upon a judgement. This will require certain concrete conditions to hold, such as an individual's membership in a society with a healthy, thick, normative framework. It will also require some internalisation of the perspectives and opportunities presented by such a society. It will also require an individual to have certain capacities, such as reflective capacities and the ability to apply abstract thought. Furthermore, ways in which an individual's level of autonomy can be expanded include:

- adopting a broader, more long-term perspective in relation to an action-decision (global-bookkeeping)
- considering a greater range of alternatives and the merits and dismerits of acting in this way ('all-things-considered' judgements)
- applying the human capacity for reflexivity (taking on higher-order judgements and mental states, which, if yielded to, result in a higher-autonomy state)

All of these means of autonomy expansion involve a broader array of materials being incorporated into decision-making (even the individual's human weaknesses and own motivational states, when it comes to higher-order attitudes), and accordingly the individual conceptualising a richer range of possibilities as available to them. The lesson seems to be

that an individual is more autonomous when they take on such a richer perspective, and that this can only be afforded to them by the presence of certain material conditions, and their possession of certain capabilities.

Some Versions of the Horizontal Account

The arguments above suggest a link between autonomy and the richness of the perspective adopted by an individual in relation to a given judgement. This lesson can be applied whether that judgement is in the form of the intention to carry out a particular action, or perhaps reflected in an attitude such as a belief or desire – a stance taken towards an object in the world (or towards another aspect of the self). Returning to the framework of the dialogical self, whichever perspective or position is temporarily adopted by the 'I' as it forms a judgement, which may be a perspective that absorbs or adjudicates between a number of other positions or perspectives, the proposal is that autonomy will be tied to the richness and breadth of that perspective. Though higher-order attitudes may have a role to play in increasing the richness and breadth of that perspective, the same criteria apply at the first-order level as at the higher-order levels, and the possession of higher-order attitudes is not used to define a given level of autonomy. This proposal suggests that autonomy is a matter of degrees, since the criteria for autonomy are not binary but rather continuous.

However, there are some choices to be made concerning the precise elucidation of the horizontal account of autonomy. So far, I have mentioned the conceptualisation of an adequate range of action options, and the cognition of a broad array of goods and facts, including abstract goods and facts. I have talked about the individual adopting a broad and rich perspective. However, for a new horizontal account of autonomy, these wide and intuition-based notions will need formalising into a more concrete set of criteria, that determine the level of autonomy of an individual. Moreover, the account needs to be formalised in such a way that respects the horizontal understanding of the self set out in the initial sections of this chapter: that the self as a whole is relevant for determining the conditions of autonomy, that the 'authoritative' self is not limited to one core element, or set of elements, and that no motivational underpinning of action is to disregarded as alien, or external, to the self. The conditions should respect the view that the relations that can hold between different aspects of the self are relations between aspects that are, in a fundamental sense, equals. In the sections below, I will outline a number of ways of tying this horizontal

conception of the self with the view that the conditions of autonomy pertain to the richness and the breadth of the perspective adopted by an agent in relation to a given action-decision, or judgement. Altogether, there are three choice points to be made concerning the structure of the horizontal account of autonomy.

Choice point (a): Assessing the Richness of the Perspective – Relevant Factors

One choice concerns the relevant factors (or set of factors) by which the richness of the perspective is to be measured. As discussed in the previous section, the horizontal account should not be denying that any motivational underpinning of action contained within the self, or any of the self's beliefs, including beliefs about reasons, is relevant material when it comes to the conditions of autonomy, and all should have claim to form part of the perspective of the authoritative self. However, in clarifying the conditions of autonomy, we require some metric by which the individual's level of autonomy is to be measured. As highlighted in the previous section, this should be a metric that indicates the richness of the perspective adopted by the agent in relation to a given judgement or action decision.

i. Action options

It may be that we stipulate that autonomy is to be tied to the individual's conceptualisation of an adequate range of action options. For the Heyman addiction example, this could translate into merely conceiving the action options as 'drink now' versus 'don't drink now' in the case of local bookkeeping, or in the case of global bookkeeping also conceiving 'don't drink now' as an instance of a wide set of actions that continue far into the future, with positive consequences. This option isn't afforded by the local perspective. For Lear's example, this could translate into the conceptualisation of action-options made available by a culture's normative framework. Membership in a shared culture affords meaningful action opportunities that would not be available otherwise. In the Heyman example, the individual's addiction is tied to their adoption of an attenuated perspective in which certain action opportunities are not available to them. In the Lear example, the relevant normative concepts that would open up meaningful action opportunities simply aren't available to be adopted into the individual's perspective.

Human reflexivity is also likely to open up a broader range of action alternatives, since it involves an individual reflecting upon their immediate motivations, assumptions, and impulses. In doing so, the individual creates space for consideration of alternatives that go beyond gratification of the currently strongest desire, for instance options that involve long term planning, or the pursuit of a broader set of possible goods than those involved in the immediate satisfaction of a desire. Therefore, reflexivity can also be understood to render further action possibilities relevant. Considering a broader range of alternatives also involves incorporating more relevant factors into the decision about what to do, since one option can only be pursued at the expense of its alternatives, and considering a wider set of possible goods to be pursued opens up more meaningful action opportunities. Therefore, this version of the horizontal account can also help explain the intuitive appeal of the notion of an 'all-things-considered' judgement, when it comes to the conditions for autonomy.

If the richness of the perspective is to be tied to the conceptualisation of an adequate range of action options, this needs to be explained in such a way that the self as a whole is relevant for securing the conditions of autonomy. However, it seems that this condition can be met. An individual's desires, beliefs including beliefs about reasons, values and commitments, are all aspects of an individual's perspective that can be determinant of the action options they perceive as available to them. Desires involve worldly objects to be attained by acting in a certain way (see e.g. Railton 2021 on desire). Beliefs may be beliefs about value, or the good, as well as factual matters that can help an individual to predict the likely outcome of their behaviour. Once again, these can contribute to the kinds of action options an individual incorporates into a decision. When it comes to an individual's values and commitments, these can directly translate into goods to pursue, which in turn open up meaningful action opportunities for an individual. Therefore, the action options metric helps explain the relevance of the self as a whole for securing the conditions of autonomy. Under this version of the account, it is not that certain aspects of the self are alien, without the potential to be autonomy-contributory. The scope of the authoritative self is not limited to one type of underpinning of action, such as beliefs about reasons, or to a limited set of attitudes of the agent, such as a limited set of their desires or a limited set of their commitments. Rather, all of these have the potential to be autonomy-contributory, under different stances and perspectives that the authoritative, holistic self can take on.

ii. Facts

Alternatively, the richness of the perspective could be formalised in terms of the breadth of relevant facts that are incorporated into the decision or judgement. This could also be expressed in terms of reasons, depending on the conception of reasons adopted, and the understanding of the mechanism by which the self responds to reasons. To avoid the pitfalls of 'hierarchical' accounts, it cannot be that this involves a distinct capacity to respond to reasons, that is separate from other possible underpinnings of action. An internal conception of reasons (see Williams 1989) could help open the scope for such a version of the horizontal account, expressed in terms of reasons. This is because an agent's reasons are defined in terms of their motivational set, helping to align the desiderative and the rational aspects of the soul.

Returning to relevant facts, in the Heyman example, the local bookkeeping individual fails to adopt relevant facts concerning the long-term impacts of their behaviour into the future, into their action decision. In the Lear example, the Crow people, following cultural devastation, simply had a more limited normative framework to adopt into their decision about what to do. In this way, there simply *were* fewer relevant facts that could be incorporated into their judgement. As noted, human reflexivity can be understood to bring a greater range of factors to bear upon a given judgement (including those that relate to the agent themselves, as a further object in the world). Furthermore, measuring the richness of the perspective by the breadth of relevant facts brought to bear clearly helps explain the intuitive pull of an 'all-things-considered' judgement as the marker of autonomy.

Similarly to the case in relation to action options, when it comes to relevant facts, it is not one limited subset of the attitudes of the self that is relevant for securing the conditions of autonomy. Attitudes are intentional states, with an object such as a feature of the world or a feature of the self, so all can have a role to play with regards to the breadth of facts that are incorporated into a judgement. They are aspects of an individual's psychology that infuse a perspective, as the individual contemplates various facts pertaining to their action-decision, or other judgement. Therefore, relevant facts can be an appropriate metric, when it comes to the conditions for the horizontal account of autonomy. Once again, the horizontal account does not operate on a model of alienation or externality, when it comes to elements of the self that can issue in action.

When it comes to the relevant factors against which the richness of the perspective is to be measured, however, whether this is in terms of action options, relevant facts, or another alternative, there is an important caveat. A purely quantitative understanding of the way in which the incorporation of relevant factors into the judgement expands the individual's level of autonomy may suggest that there is no end to the amount of information that would ideally be incorporated into a judgement, in order for an individual to reach the highest possible level of autonomy. This could suggest that an outcome of practical paralysis, in which an individual attempts to incorporate more and more information and continually questions any judgement they form, is the recommendation of the horizontal account of the self-relation involved in autonomy.

There are a number of ways, however, that this challenge can be met. One is by alternative understandings of the relationship that must hold between the agent and the relevant factors, in order for those factors to be determinant of the agent's level of autonomy: this I will expand on in choice point (c) – relation to agent. Another is by limiting the scope of factors that are deemed *relevant* to the given judgement – if factors beyond these are considered, then they will not be contributory to the agent's level of autonomy. The final option is to stipulate that a broad array of factors can only be autonomy-contributory, to the extent that it can successfully be incorporated into a decision or judgement, and that the appropriate amount of time to apply to deliberation is a further choice point for the agent, that can also be carried out more or less autonomously. This judgement should factor in our practical limits as deliberative beings, so that we do not end up with the perverse result that the most supremely autonomous agent will deliberate endlessly.

Before moving on to choice point (b), I will discuss a further possible objection to the formulations of the horizontal account that have been detailed so far. Rather than being tied to a broad and expansive perspective, there may seem to be cases in which an attenuated perspective is in fact autonomy-promoting. Recall my discussion of Frankfurt's notion of volitional necessities, in chapter two (Frankfurt 2002, 2006). Frankfurt suggests that when we identify with an aspect of our psychology, for instance the love we hold for someone, this sets limits on what it is possible for us to will. This seems to in fact reduce the scope of possibilities available to one, and to reduce the range of facts or goods that are relevant in deciding what to do – love can be capable of crowding out other considerations, and closing off the possibilities of our pursuing other things. For instance, if we love someone, we might find ourselves unable to leave them, even if that is severely restrictive upon our available

action options. A parent, for instance, might find themselves with fewer action options available upon having a child. However, when a person acts in pursuit of core commitments like these, this is autonomy-contributory as opposed to autonomy-undermining – the pursuit of such commitments is a pattern of behaviour that reflects themselves in the deepest and truest way.

However, there is another way of construing an individual's relationship to such commitments that supports the idea that autonomy is supported by a rich and broad perspective. When it comes to love, for instance the love a parent holds for a child, no matter how much information the parent assesses, and no matter how many alternatives the parent considers in terms of the action-options available to them, this will all only be supportive of their judgement that the thing to be done is to pursue their child's best interests. The broadest possible assessment will only support the conclusion that this is the way to behave. This is in contrast with, for instance, addiction behaviours, in which the conclusion that the thing to be done is to take the drug now, can only be supported by the application of a limited perspective that overlooks a large number of better alternatives and a large number of considerations against carrying out that thing. I therefore contend that the kind of crowding of alternatives that is implied by an individual's deepest commitments is of a different kind to the autonomy-undermining overlooking of alternatives involved in low-autonomy states, such as those of a relapsing addict. Even if a parent does not always consciously consider all the alternative action options to caring for their child, doing so would produce the same conclusion (see choice point (c) for more about the possible relation between the agent and the relevant factors).

Choice point (b): Benchmark

Moreover, there is the choice point of the benchmark which the relevant factors are to be assessed in relation to. It might be that we regard the agent as being more autonomous, the greater the range of factors afforded by the world, that are brought to bear upon a given judgement. Perhaps the more action options afforded by the world that are incorporated, the more autonomous the individual, or the greater the volume of worldly facts that are incorporated, the more autonomous the individual. Alternatively, the benchmark by which the relevant factors are to be assessed might be by reference to the psychology of the individual. Perhaps autonomy is merely to be defined in terms of action-options that the individual is

already on some level aware of, or facts or reasons that the individual is already on some level aware of. In this case, the individual would be more autonomous the greater the range of relevant materials already contained within the self, that are incorporated into the judgement.

The former sort of benchmark would suggest that autonomy can be increased by knowledge and education, and is better able to account for the kind of example contained within Lear's *Radical Hope*, in which the conditions for autonomy seem to depend on the agent's relationship with a world, that can afford them certain opportunities. Once deprived of the old normative structure around which to build their lives, individuals within the Crow nation might have still made best use of the remaining normative concepts that were available to them, and absorbed by them, in the circumstances.

On the other hand, treating the benchmark as the self as a whole aligns well with the observations around human reflexivity and all-things-considered judgements, since the inclusion of higher-order attitudes increases the scope of the self that is incorporated, and an individual can only apply factors that they are aware of into an 'all-things-considered' judgement. It can help explain why action that is autonomous is understood to issue from the agent in a deeper and a truer sense than action that is not autonomous. Moreover, this kind of benchmark helps to clearly explain the role of the state of the self as a whole for determining the conditions of autonomy – the more of the relevant factors contained within the self that are incorporated into a given judgement, the more autonomous the individual. It also opens the possibility of alternative relevant factors that the horizontal account might adopt when the benchmark is their relation to the self as a whole.

When the benchmark for autonomy is the extent of relevant material contained within the self that is incorporated into decision-making, we open the possibility of defining the conditions for autonomy in terms of an action or a judgement's relation to attitudes of the self as a whole, such as its entire set of desires, or its entire set of values. We might say that an individual is more autonomous, the greater the range of their desires that are incorporated into the action-decision. Alternatively, we might say that an individual is more autonomous, the greater the range of their values that are incorporated into the action-decision. This provides a simple and intuitive picture of the way in which autonomy is to be tied to the condition of the self as a whole.

Applying a metric of desires or values may seem to limit the types of aspects of the self that are relevant for securing the conditions of autonomy. However, these are types of attitudes

that can underpin both action that involves the most attenuated perspectives, and exhibits the lowest level of autonomy, and action that exhibits the widest and richest perspectives, and greatest levels of autonomy. Moreover, in each of these types of states, from e.g. states that incorporate a limited number of desires to states that incorporate a large number of desires, these attitudes will interact with other aspects of the self, such as their beliefs and commitments, in a way that infuses and informs the relevant perspective. Once again, an internal conception of reasons (Williams 1989) may have a role to play in explaining the interaction and integration between different types of aspects of the self.

However, even when the benchmark is the world as opposed to the self, it is still possible to explain how the self as a whole is relevant for determining the conditions of autonomy. In contrast to hierarchical accounts, that isolate one limited aspect, or type of aspect, of the self, that has a role to play when it comes to the conditions of autonomy, the horizontal view does not limit the authoritative self to one core and permanent element. When a greater range of action options are considered, or greater range of facts in the world brought to bear upon a given judgement, the self that is carrying out that judgement needn't be defined in one limited and permanent way. Returning to the model of the dialogical self, it might be that the self-as-knower adopts one of a number of different perspectives or positions, but it is not to say that it is autonomous in one, whereas it fails to be autonomous in another – it can be autonomous to different extents in each of the separate positions. It might also take up a position in relation to these various positions, as a result of the conflict between the various I-positions. In this case, there is another perspective that can be assessed against the worldly autonomy benchmark. However, none of these positions represents the final, authoritative self, and autonomy is always a notion admitting of degrees.

Choice point (c): Relation to agent

As well as the type of factors deemed relevant for assessing the richness of the perspective, and the benchmark that they are assessed in relation to, there is a further choice point for the horizontal theory of autonomy: the type of relation that needs to hold between the individual and this range of factors, in order for them to be determinant of the agent's level of autonomy. The type of language I have been using so far has mainly implied that these factors need to be *consciously* incorporated into an action-decision, or, more generally, into a given judgement. The language of 'perspective' suggests an outlook that the individual is

consciously aware of, and it might seem that horizontal autonomy is naturally tied to the range of action-options, facts, desires, or values, that are consciously incorporated into a given judgement in the process of deliberation. Under this proposal, autonomy is executed through a transparent process that the individual is aware of. This fits well with the conception of the 'self-as-knower' suggested by the dialogical conception of the self. However, this is not the only option.

There is reason to regard conscious consideration of action-options etc. as too strong a requirement for those factors to count as *incorporated* into the decision or judgement. In relation to rationality as opposed to autonomy, but discussing considerations that have relevance here, Arpaly (2000) suggests that an individual can sometimes be *more* rational for acting against their better judgement than for acting in accordance with their better judgement. This is when reasons cause an individual to behave in a certain way, via a mechanism that bypasses their deliberative processes, yet the outcome is more rational than the course of action decided upon by their best judgement. One of her examples is a student panicking about their final exams, and deciding to become a hermit and spend all of their time studying. However, such a choice is ultimately not rational for them, since it is so detrimental to their happiness, and they care about their happiness above all else. In fact, their concern for other aspects of their life means that they do not achieve their goal of becoming a hermit, but this is a more rational outcome than the course of action recommended by their best judgement. Altogether, one's best judgement is not infallible, and there may be cases in which we act better through processes that bypass our best judgement.

Applying these considerations to the horizontal theory of autonomy, perhaps it should not be a matter of the range of factors that are consciously considered, but rather the range of factors that have a causal effect upon the judgement (potentially with an 'in the right way' stipulation, to avoid instances of wayward causation). Alternatively, we might merely consider the range of relevant factors that are in fact satisfied by the judgement. If our benchmark is the self, this latter option would translate into the desires or values of the agent that are satisfied by the judgement, even if they are not consciously considered. Alternatively, if the world is the relevant benchmark, it would translate into the judgement's relation to the world as a whole (in terms of its relation to the action-options available, or the relevant facts or reasons, etc.), bypassing consideration of the agent's deliberative processes.

Finally, the relation between the agent and the relevant factors might be expressed in terms of those that are available or salient to the individual. Rather than considering the breadth of factors that are consciously incorporated into the judgement, or that have another kind of causal impact upon the judgement or are merely satisfied by the judgement, we might consider the range of factors that are available or salient to the individual, to be incorporated into their judgement. It might be that we want to hold that if, for instance, an action-option was available to an individual, they could have incorporated it into their decision-making, even if they did not. In that sense, we might say that they acted based on a limited set of material, but not in a way that was autonomy-undermining. This would however require clarification of the precise application of the notions of 'availability' and 'salience' that we wish to employ.

For example, availability might require membership in a common culture with the required normative concepts, for an adequate range of action options or an adequate range of normative facts to be incorporated into decision-making. Availability might involve possessing certain skills and capacities, that enable the individual to grasp the relevant concepts. We might even go further and say that for those facts or concepts to be available, the individual must know or be aware of them.

With regards to salience, we might adopt a definition like Yetter-Chappell and Yetter Chappell's (2016, p449) copied below:

> 'A feature of the world is salient to an agent t at a time t in so far as that feature exerts an involuntary draw on the individual's attention at t'

For Yetter-Chappell and Yetter Chappell, salient things are those that grab our attention. We could apply this to the construal of horizontal autonomy and state that an individual is more autonomous, the greater the range of relevant factors that are salient to the individual. By definition, salience is beyond (at least direct) voluntary control, so which factors are salient to the individual is a plausible candidate for determining an individual's level of autonomy. If relevant factors are not salient to an individual, this could be intuitively autonomy-undermining.

Altogether, these three choice points for the horizontal account of autonomy represent three parameters which would need to be filled out by a fully-fledged conception of the horizontal theory of autonomy. In the table below, each letter represents a possible form that the horizontal theory of autonomy could take. Whilst I have alluded to some advantages and

disadvantages of the various options above, it is beyond the scope of this thesis to make a final case for one option in particular.

Table 1: Possible Construals of the Horizontal Account of Autonomy

Benchmark	Relevant Factors	Relation to Agent			
		Consciously Incorporated	Merely Satisfied	Available	Salient
The World	Action-Options	a	b	c	d
	Facts	e	f	g	h
The Self	Action-Options	i	j	k	l
	Facts	m	n	o	p
	Desires	q	r	s	t
	Values	u	v	w	x

<u>Conclusion</u>

In conclusion, I propose that we level the hierarchy implied by hierarchical accounts of personal autonomy and move towards a horizontal account. Hierarchical accounts suggest that the authoritative self is to be limited to one distinct aspect, or set of such aspects, but I argue that the self as a whole has relevance for determining the conditions of autonomy. Autonomy is not merely to be achieved through the influence of one limited aspect of the self, or set of limited aspects of the self. Rather, the self is capable of adopting multiple different perspectives or positions, and no single one can have claim to be the final, ultimately authoritative self.

However, higher-order mental states and 'all-things-considered' judgements have intuitive appeal as the criteria for autonomy. It seems that this, together with specific examples of autonomy-deprivation described by Lear (2008) and Heyman (2010), can be explained by the

richness and breadth of the perspective adopted by an agent when it comes to a particular judgement, whether that is a judgement in the form of an action-decision, or in the form of an attitude taken towards something. The specific elucidation of a horizontal account of autonomy would require choosing the values to be taken by three choice parameters: the relevant factors by which to judge the richness of the perspective, the benchmark which those factors are to be assessed in relation to, and the kind of relation that is to hold between the agent and this set of factors.

In the next chapter, I will make the case for this form of account of personal autonomy, with a positive argument in favour of it, together with a demonstration of the ways in which it can surmount the challenges to hierarchical accounts I discussed in the previous chapter.

Chapter 4: Defence of the 'Horizontal' Conception of Personal Autonomy

Introduction

In this chapter, I will make the case for moving away from 'hierarchical' accounts of personal autonomy, towards a 'horizontal' alternative, the possible forms of which I set out in the previous chapter.

First, I will set out two related positive arguments in favour of the conception of the self involved in the 'horizontal' account of autonomy, as set out in the previous chapter. Then, I will outline the ways in which this type of alternative is able to surmount the criticisms I have applied to 'hierarchical' accounts, as set out in chapter two.

These considerations, together with those already set out in the previous chapter, suggest that we should be moving away from the dominant form of account, the 'hierarchical' account, towards a 'horizontal' view of the self-relation involved in personal autonomy.

The Horizontal Account and the Structure of the Self

In support of the model of the self set out in the previous chapter, I argue that the human capacity for reflexivity is limited. Whenever we self-govern, or take a stance in relation to an aspect of ourselves, or an aspect of the world, we can only do so on the basis of certain motives or beliefs about reasons which we *apply*, whilst refraining from judging or bringing into question. This means that whichever perspective we adopt always involves elements that are the *materials* of scrutiny, as opposed to the *objects* of scrutiny. They are the background tools that infuse a perspective, and inform a judgement, without being brought into question themselves. This means that our forming a judgement, our taking a stance in relation to aspects of ourselves, and aspects of the world, can only ever be a partial affair. It involves withholding judgement on aspects of ourselves, attitudes such as beliefs and desires that inform our perspective, whilst they are used to assess other aspects of ourselves. Moreover, these 'tools' that are employed in taking on such a perspective can always be the potential objects of scrutiny, from other perspectives that we might employ.

These considerations support both the idea that an 'all-things-considered' judgement can never truly be all-things-considered, and that human reflexivity, or higher-order attitudes, do

not bring us to a point which can be legitimately labelled as the 'top' of a hierarchy, or the representative of the final, true self. Even an nth-order higher-order attitude will involve the application of a limited capacity of reflexivity, and will involve employing 'tools of scrutiny' which can themselves be the object of scrutiny, if the self takes on another stance. This points in favour of my suggestion in the previous chapter that we 'level the hierarchy', in line with a horizontal conception of the self and personal autonomy. It supports the notion that the conflicts and relations between different aspects of the self are between aspects which are ultimately, in a sense, equal, and none of which have ultimate or final claim to be the 'true', authoritative self.

This also runs counter to hierarchical accounts of autonomy in general, since they are characterised by limiting the purview of the authoritative self to one distinct element, or set of elements. In fact, whichever element is chosen, or whichever position is chosen as the 'final' or 'ultimate' representative of the agent, it involves the application of materials on which judgement is withheld, but which are the possible objects of scrutiny, from another point of view. The potential processes of self-governance are never final, and attempting to ground authority in one limited aspect or one limited set of aspects of the self will ultimately be unstable. Whilst there is no demand for us to continually reassess our own perspectives, the possibility is always there, and open to us. This undermines the notion that whichever single 'authoritative' aspect is chosen truly represents the agent's final point of view on a matter.

I will highlight two related arguments in support of this picture. One, inspired by Ware's (2015) remarks on Fichte, appeals to motivation. The second concerns a third-person perspective which we apply to others, but can also relate back to ourselves. To help support my second argument, I will appeal to Moran's (2001) distinction between theoretical and deliberative stances.

Ware (2015) argues for an interpretation of Fichte's *System of Ethics* (SE) that employs the following principle, which presents a challenge to contemporary Kantian theories of normativity (Ware 2015, p3):

> *Limited Reflective Authority*. We cannot suspend the defining features of ourselves – our inner drives or maxims of choice – to the point where they appear normatively arbitrary.

This is in contrast to (Kantian) views which hold that it is possible for us to attain full reflective authority – that as rational beings we can entirely distance ourselves from our

motives and assess them with reference to reasons (for instance as implied by Korsgaard (1996)). In contrast, under limited reflective authority, whenever we deliberate, our motives, or inner drives and maxims of choice, must retain their force to some extent. It is not the case that we can ever entirely step back from our motives and coolly assess them. The deliberative self will always still be operating under the influence of the normative significance of our inner drives or motives. Under Ware's interpretation of Fichte, this limited reflective authority contributes to the processes of moral progress. The stages of moral progress are driven by our ability to reflect upon our natural drives, for instance when we reflect upon the drive of self-preservation (drive that characterises stage 1), this opens up the possibility of a new maxim – that of happiness, giving us the capacity to delay gratification and plan our levels of satisfaction over time (stage 2). However, it is never the case that in reflection we are entirely distanced from our underlying drives and motives. For instance, the reflection that underlies the progress from stage 1 to stage 2 is only a matter of questioning the ways in which my initial natural drive can be fulfilled, as opposed to whether it should be adhered to. At stage 3 the agent is brought into awareness of the highest drive, that of self-sufficiency, but even when they fully contemplate this drive at stage 4, the final stage, they do not operate outside the force of their natural drives, but rather their deliberation is infused by them.

Ware (2015) argues that this thesis of limited reflective authority can help to counter a sceptical challenge that can be applied to modern Kantian theories: their inability to explain our motivation to be moral, or to follow the laws of reflection or reason. Ware (2015) suggests that this problem only arises if we assume full reflective authority – that it is possible to distance ourselves from all of our underlying drives and subject them to the scrutiny of reasons. If a modern Kantian maintains that to be moral is to follow the laws that are constitutive of our rationality, or our agency, then we can still be left with the question 'Why be an agent?' If we are entirely distanced from all of our motivations, then there is no answer to this question. On the other hand, the thesis of limited reflective authority closes off the conceptual space in which a question like 'Why be an agent?' can gain traction – there is no space to ask why we should be bothered about being agents, since our motives retain their force from the point of view of deliberation. Limited reflective authority ensures that motivating elements are always contained within the deliberative process itself, thus filling in the gap created by accounts of full reflective authority.

I argue that this thesis of limited reflective authority is the correct way to understand the self, and that this buttresses the argument that the processes of self-governance are only ever

partial, and no one limited aspect of the self can legitimately stake a claim as the final, authoritative aspect. We do not need to support a Kantian conception of agency and / or morality, to apply similar considerations around motivation to those highlighted by Ware that can support this view: the reflecting self, that takes a stance or makes a judgement in relation to something in the world, or an aspect of the self such as a drive, cannot do so in a manner that employs full reflective authority. If it did so, then it could have no interest in the results, and no motivation to employ such deliberative capacities or form such a judgement in the first place. Whether the self is assessing a given desire with reference to reasons, or assessing reasons in relation to other reasons, there must be certain drives or normative inclinations within the self that are applied, as opposed to being brought into (at least full) scrutiny. The self can never distance itself from itself completely – this can only ever be a partial matter.

This also means that hierarchical accounts in general fall prey to the kinds of regresses that have been discussed in relation to Frankfurt's (1971) higher-order volitions account of autonomy, and hierarchical accounts are ultimately unstable if they claim to have isolated the relevant core, authoritative, final aspect of the self. Whichever outlook is said to represent the outlook of the 'authoritative' self, it will involve elements which are the materials of scrutiny, as opposed to the object of scrutiny. It can never therefore present the final view on a matter. It will always involve elements which can be scrutinised from the perspective of other elements of the self. This supports the dialogical and horizontal conception of the self as set out in the previous chapter, and the notion that the self as a whole, as opposed to one limited aspect or one limited set of aspects of the self, has relevance when it comes to securing the conditions of autonomy.

This can also be defended by considering what goes on when an individual makes a given action-decision or forms a given judgement, from a third-person perspective, and applying those considerations back to ourselves. This holds even if the given action-decision or judgement is in the form of a second or third-order volition, to use Frankfurt's terminology (whether or not that is a *decisive* higher-order volition), or the reaching of a commitment or stance of the kind that might be said to compose an individual's 'character', under a character account of autonomy (see e.g. Ekstrom 1993), or the host of other possible manifestations. When other people make decisions, or reach a view about various things, such as an opinion on a matter, we understand them as doing so on the basis of some kinds of materials. Our explanations of their behaviour might take the form of rationalising explanations, in which we explain their behaviour by reference to reasons that justify the way they have behaved.

Individuals might also put forward such explanations in relation to themselves, whether spoken to others or thought in relation to themselves. Alternatively, our explanations of behaviour might take the form of causal explanations, in which we explain actions in a similar kind of manner to the way in which we explain other happenings in the world – cause and effect. For instance, we might claim that a particular desire or disposition was the cause of an individual pursuing a destructive course of action.

When we do so, we grant that the given judgement or action-decision formed by the individual, is not the final element of the explanation of their behaviour. There is a further story about what goes on here, that appeals to aspects of that individual's psychology (whether we are applying the rationalising explanation, or rather applying the casual explanation), or alternatively appeals to factors that have led that individual's psychology to be the way it is. However, if we grant that this kind of explanation can be applied to others, we also must grant that this kind of explanation can be applied back to ourselves. In the same way in which we observe others' behaviours and judgements, and come up with rationalising and causal explanations, other people can observe us and theorise about us in similar ways. If such explanations are applicable to others, then such explanations are also applicable to ourselves.

However, when it comes to other people, these explanations are applicable even when another person is acting on the basis of a value that is constituent of their character, or responding to their understanding of the relevant reasons, or forming what they take to be an 'all-things-considered' judgement about an attitude or course of action. Similarly, when we take on such a stance ourselves, this is not the final element in the explanation. Once again, such stances indicate the presence of underlying materials, such as factors like desires and tendencies that have caused us to take a particular stance, or reasons that we have incorporated into our judgement, whether consciously aware of them or not. This implies that there are further possible objects of reflection that are not assessed as we form the given judgement, but are rather the tools by which that judgement is formed. There is always the potential for us to reflect upon the sources of our judgements and behaviour, and the reasons that have informed them. Once again, therefore, this indicates that the relevant aspect of the self said to represent the 'true self' under hierarchical accounts, does not have convincing claim to present an agent's final point of view.

Moran's (2001, chapter 2) remarks on deliberative and theoretical stances have application here. Moran argues that there are two different possible, and interacting, stances which we can take towards our own mental states and behaviour. We can take on a theoretical stance, in which we assess our attitudes and choices in a similar way to that in which we assess those of others. We consider the causes that have led us to behave in a certain way, or the reasons that we have taken to justify a given stance or course of action. We might predict what it is likely that we will do based on our tendencies or predispositions. However, this kind of theoretical stance is not the normal relation we have to our mental attitudes, our beliefs and intentions and pro-attitudes. Instead, when someone asks us what we are to do, for instance, our outlook is outwards, towards the reasons for and against carrying out a course of behaviour. This is the deliberative stance that is our normal relation to our attitudes, our mental states that are directed towards the world. That our normal relation to our attitudes is the deliberative stance might create the impression that we can settle on a point of view that can represent the outlook of the final, authoritative self, for instance in the form of an attitude resembling Frankfurt's decisive higher-order volitions. However, the possibility of taking up the theoretical stance always remains there. Also, as Moran notes (chapter 1), the special form of immediate access we can have to our first-person mental states in the form of the deliberative stance does not undermine the substantiality of that knowledge. These first-person mental states are still pieces of the world, with a history and a story that can be the potential object of scrutiny, in a similar manner to the way in which we assess the mental states of others. Once again, these reflections indicate that whichever 'authoritative' aspect of the self is chosen by a hierarchical account, it cannot have final claim to be the sole, true authoritative self.

Returning to Criticism (1) of 'Hierarchical' Accounts: Explaining the Possibility of Nonautonomous Action

In chapter two, I criticised 'hierarchical' accounts for failing to explain the possibility of nonautonomous action. They fail to adequately explain how the authoritative self would allow itself to be sequestered by foreign powers, or how, if this did indeed occur, the resulting behaviour could truly be an action ascribed to the self as a whole, never mind an intentional action ascribed to the self as a whole. This runs counter to apparent and convincing examples of nonautonomous action, and nonautonomous intentional action. These include goal-directed behaviour that an agent might carry out to further an addiction, or

individuals whose autonomy is undermined by a minimal set of options to choose from, even as they intentionally direct themselves within that limiting context.

Since 'hierarchical' accounts identify one limited aspect of the self, or one limited set of aspects of the self each with a similar structure, as the 'authoritative' self, that aspect or those aspects of the self seem to have sole claim to represent an agent's point of view on a situation. Therefore, when an agent is moved to behave by motivational impulses that run counter to those aspects of the self, it is not convincing that the resulting behaviour can be labelled as the agent's own, even in the minimal sense for it to be labelled as the agent's own action (in contrast to the agent's own autonomous action).

For a 'horizontal' account, on the other hand, the self as a whole has relevance when it comes to determining an agent's level of autonomy. There is not a single permanent, or set of such permanent, positions that represent the perspective of the 'authoritative' self, and low-autonomy states involve features that can also be relevant for high-autonomy states. Moreover, the path from nonautonomy to autonomy is a continuous path, admitting of degrees, as opposed to the crossing of a binary nonautonomy / autonomy distinction. Even when an agent is in a high-autonomy state, this is a condition at a point on a spectrum, and the same kinds of features that lead to low-autonomy states will still be in place, only in a reduced sense. There is no sense in which an agent can be 'completely' autonomous. This gradated understanding of autonomy also opens the possibility of nuance and new forms of explanation when it comes to understanding how failures of the authority relation could come to occur.

As discussed within the previous chapter, a 'horizontal' account allows each of the distinct aspects of the self, and each set of impulses and motivating views about reasons, to offer itself as part of the perspective of the 'authoritative' agent. It is not that an agent is ever entirely alienated from any of its parts – all of them have claim to be part of a holistic, wide interpretation of the 'true self'. Take Frankfurt's (1971) unwilling addict, who has a first-order desire to take the drug, a second-order desire not to take the drug, and a second-order volition that the desire not to take the drug is the effective one in leading the individual to action. Whether the unwilling addict is ultimately moved to action on the basis of their first-order desire to take the drug, or their first-order desire not to take the drug, when the case is construed under a horizontal conception of personal autonomy, each piece of behaviour can have convincing claim to be the agent's own action, and even their own intentional action.

Both desires form part of the 'true' self, as widely construed under the horizontal account. Returning to the model of the dialogical self, the self is able to take on different perspectives, such as the perspective of the desire to take the drug and the perspective of the desire not to take the drug. It is not to say that one perspective, or one set of perspectives, has sole claim to be the 'true' self. Therefore, when an agent acts on one impulse in a manner that we deem to be nonautonomous, such as the unwilling addict succumbing to their desire for the drug, we do not need to go so far as to deny that this is their action, or their intentional action, also.

Rather, the unwilling addict's level of autonomy, as they act on the basis of their desire to take the drug, or their desire not to take the drug, is tied to the attenuation or the breadth of the perspective adopted as they implement their action decision. When they act on their desire to take the drug, this can only be by overlooking a vast range of considerations against carrying out that thing, and a vast and rich range of alternative action options that could be available to them. Whether our benchmark for autonomy is the action judgement's relation to relevant factors contained within the world, or whether our benchmark for autonomy is the action's relation to the self as a whole, the unwilling addict's action on the basis on their desire to take the drug involves incorporating a limited view in relation to the situation at hand. In this way they exhibit a low autonomy state, without the need to deny that this is their own action, and their own intentional action.

They act in a way that exhibits low autonomy, but on the basis of a point of view informed by beliefs and impulses that the horizontal account does not deny form part of the authoritative, holistic self. These beliefs and impulses can also be autonomy-conferring, when adopted into wider and more encompassing perspectives. In implementing the position that they do, they are not taken over by some alien aspect of the self. The desire to take the drug forms part of a wide-ranging and varied self, and is relevant material that could be incorporated into other perspectives which the self might take on, even perspectives that offer themselves up as supremely high-autonomy states. The perspective they adopt is simply at the lower end of a continuum, pertaining to the breadth of their perspective, and this continuum will lead right up to the position of the most supremely autonomous agents. In this way, we do not deny that the action is the agent's own, though we offer an explanation for why the unwilling addict who acts on the basis of his desire for the drug is in a position of low autonomy, as opposed to high autonomy.

In contrast, the unwilling addict who acts on the basis of his desire not to take the drug can be understood to adopt a wider and richer perspective in relation to the situation at hand. His action-judgement reflects a wider set of his commitments and desires, and better reflects all the relevant information and opportunities that the world could provide, in relation to his action decision. However, this can never be complete. As mentioned in the previous chapter, the notion of an 'all-things-considered' judgement presents a conceptual impossibility. There will always be shortfalls and oversights, even in a high-autonomy state. Therefore, the kinds of aspects of an agent's psychology that are autonomy-undermining are still present in a high-autonomy state, only to a lesser degree. This helps to support the contention that behaviour can still be the agent's own even when it issues from a low autonomy state, in which these features are present to a greater degree. There is no discrete divide between nonautonomy and autonomy, but rather there is an agent who does not exhibit fundamentally different characteristics in each distinct state. Whether the agent is autonomous or nonautonomous, it is still possible for the agent to bring about action that can be ascribed to it as a whole.

Moreover, it becomes easier to understand how the authoritative self can allow itself to fall prey to processes that lead the agent to act in a nonautonomous manner. It is not that some foreign force manages to interject and hijack the processes of the authoritative representative of the self. Rather, the processes that occur are internal to the self, understood as a wide and holistic whole. As discussed in the previous chapter, low autonomy involves the self overlooking certain things, or failing to incorporate certain things, as it applies certain other things, that are still internal to the self. Take the case of the unwilling addict who acts on the basis of their desire for the drug. The desire for the drug is still part of the true self, it is just that many other aspects of the true self and many other relevant considerations are overlooked as it is applied when it is implemented in action. This is natural and even occurs, though to a more limited degree, in high autonomy states. Moreover, since the process is not fundamentally different to the kind of processes that occur in high autonomy states, it is easier to understand how a self could allow itself to get into such a state. In the given moment, it might even be possible for the self to apply self-deception, and convince itself that there are no other relevant considerations, that would count against performing this as the best action. There are blind-spots involved in any judgement, and the kinds of blind-spots involved in a low-autonomy state may be confused with the kinds of blindspots involved in high-autonomy states. The kinds of attenuated perspectives involved in low autonomy states

might not be experienced by the agent as particularly attenuated perspectives, especially since a perspective will always be attenuated to some degree, even in a high autonomy state.

Tenenbaum's (2020) remarks about instrumental rationality (as well as Heyman's remarks around addiction involving myopic choice, as explained in the previous chapter) provide further support towards this picture of intentional but nonautonomous action. Tenenbaum (2020) rejects conceptions of instrumental rationality that involve momentary choices and determinate results as to the course of action to take. He argues instead that human decision-making is generally carried out on the basis of long-term, indeterminate ends, that will not necessarily determine one unique option to pursue at the moment of action. Tenenbaum (1.1) discusses the example of his subway ride to work, in which he just sat and daydreamed, as opposed to a multitude of other options that would seem to bring about additional benefits, such as reading a student paper or the newspaper. However, to argue that he failed to carry out what he most wanted or preferred to do seems to assume two things, neither of which seems particularly convincing. One is that he *possesses* such determinate preferences, that give a unique answer to what is the best thing to do in any particular situation. This seems quite implausible, considering the vast array of things we could do in any situation and the fact that it might not always be so important to us what we do in a particular situation, as the subway example indicates. The second assumption is that an instrumentally rational agent will always choose the most preferred option. Tenenbaum argues that this is implausible, mainly because human beings are generally engaged in actions that extend over periods of time, of unspecified length (e.g. the subway journey), and in order to pursue ends that are not fully determinate (e.g. there is a range of ways in which the end of travelling to work may be realised, and some realisations of this end may be better than others, and some ways of realising the end may realise other ends to better and worse degrees, etc.)

Tenenbaum's account of instrumental rationality suggests that the rationality of pursuing ends across time cannot be reduced to the rationality of the choices that an individual makes at each point in time in the interval (non-supervenience thesis, chapter 2, p48). Therefore, an individual could make a rational choice at each moment between time 1 and time 2, but turn out to have been irrational over time 1 to time 2, say by failing to achieve an end of theirs. It might be, for instance, that the addict intentionally injecting themselves with drugs is the rational choice at a particular time, because the degree of pain-cessation that could be gained is so great, and they would only have been able to make an incremental contribution towards their work or family commitments within that space of time. However, this short-term

rational behaviour at individual points in time might translate to irrational behaviour over time, as they ultimately fail to fulfil commitments that are important to them. It could alternatively be the case that at this specific moment in time, their long-term preferences are not enough to determine a unique answer to what is the short-term rational thing to pursue, whether that is temporarily providing cessation to their intense pain or making some sort of contribution towards their long-term commitments. By conceiving an individual's pursuit of goals as involving long-term preferences and indeterminate ends, we can help open up conceptual space in which intentional but nonautonomous action might lie.

Such considerations indicate the way in which nonautonomous action can result from 'blind-spots' in an individual's judgement, but blind-spots that are in fact integral to the normal functioning of human deliberative capacities. It is inevitable that an individual will not fully rank all the options that are available to them, and will not possess firm preference-orderings that determine one correct way to act in a situation. This is a crucial feature of human behaviour, human action, and human intentional action, in general. However, low-autonomy states can then occur when these blind-spots mean that an individual does not apply a long-term or broad enough view in relation to a situation. This could be when short-term rational judgements, and the accompanying intentional actions, do not translate into long-term autonomous behaviour, when (referring back to Heyman), the perspective is limited to individual actions rather than continuing instances of such act-types over time. Moreover, given the structure and the gaps within an individual's preference-ordering, the individual may rationally and reasonably pursue a certain end, but since they do so repeatedly, this is at the expense of other ends that are more important to them – once again, the low-autonomy state, with an attenuated perspective, can reasonably be understood to issue in intentional action, despite failing to fulfil the conditions for a given level of autonomy. In these ways, the horizontal account can help address Buss' (1994) challenge of explaining nonautonomous but intentional action.

'Hierarchical' accounts, on the other hand, cannot be aided by these explanations in the same kind of way. For such accounts, the relevant core features of the self represent the agent's own point of view on a situation. This is not a notion admitting of degrees, and does not depend on the extent of relevant material that goes into the authoritative perspective (except perhaps, with the exception of certain variations of reasons-responsiveness accounts, though the relevant factor here is the piece of behaviour's alignment with what reason in fact depends, or what the individual takes reason to demand – if this condition is not fulfilled,

then the 'action' is not the agent's own, simpliciter.) For these accounts, nonautonomous action rather occurs when contrary influences bypass the agent's own view on a situation. Accounts in particular like Raz's, that concern reasons-responsiveness, and accounts like Frankfurt's, that involves endorsement, seem to suggest that there is one correct way, or a limited set of correct ways, for an agent to act in a situation – those represented by the correct response to the relevant reasons, or by the course of action that an individual endorses. However, this then does not allow the kind of nuance involved in an account that recognises that autonomy is on a spectrum, and that the kinds of preferences involved in judgements are not necessarily complete. In contrast, as detailed above, the horizontal account allows space for new kinds of explanations of intentional but nonautonomous action. Under the horizontal account, features that are present in high-autonomy states, where an action is indisputably an agent's own, in fact also contribute to low-autonomy states.

Returning to Criticism (2) of 'Hierarchical' Accounts: Integration

As well as failing to explain the possibility of nonautonomous action, in chapter 2 I also discussed the way in which 'hierarchical' accounts fail to explain the mechanism by which the governing self is able to wield authority over the governed aspects of the self, in the case in which authority is successfully wielded and the agent acts autonomously. In order to explain failures of autonomy, 'hierarchical' accounts must grant that motivational forces that run contrary to the dictates of the authoritative self have the power to move an individual to act or to behave in a certain way. However, given the disparate ways in which hierarchical accounts define the self-as-governed and the self-as-governor, it is not then clear how motivational forces that can operate independently of the authoritative self are also able to fall under some kind of governance relation, when it comes to successful self-governance. In the normal case, an agent acts autonomously, and authority is successfully wielded. However, hierarchical accounts struggle to explain how, given that a self can come apart in the manner implied by failures of autonomy, the authoritative self can have traction over the rest of the self as a whole.

The horizontal account of autonomy, on the other hand, questions the validity of this entire picture. It is not the case that one limited aspect of the self is the 'true self', or that there is a self-as-governed and a self-as-governor that are defined in fundamentally different ways. On the contrary, the self as a whole is the true self, and the relations between different aspects of

the self are relations between aspects that are fundamentally equals. The dialogical conception of the self, as applied to the horizontal account, highlights the way in which the I, or self-as-governor, can take on different perspectives, that absorb different aspects of the self, at different points in time. Such perspectives may involve the self taking a position in relation to different features of the world, such as forming a judgement about a matter of fact, or forming an action-intention. In doing so, it will be operating from a perspective that will be more or less wide, and accordingly will exhibit a higher or a lower level of autonomy.

Alternatively, such perspectives may involve the self taking a position in relation to different features of the self, as it employs its reflexive capacities. This is more similar to the model of self-governance implied by hierarchical accounts, which may for instance involve the self as rational being taking a position in relation to a given desire. However, for the horizontal account, whichever feature of the self is the object of reflection in such an instance, it also has equal claim to be part of the authoritative self, and at another point in time it might form part of an alternative perspective that the self takes on in its capacity as self-as-governor. It might even form part of the stance of the agent in their most supremely autonomous capacity. Higher levels of autonomy involve an agent taking on a wider view in relation to a given judgement, or action decision. However, this is always still from a partial perspective, and aspects of the self not contained in that perspective always have a potential role to play, when it comes to the autonomous individual – no aspect is irrelevant to the conditions for autonomy, nor falls out of the possible scope of the self-as-governor.

The I or self-as-governor may take on the position of various reasons against, say, carrying out a course of action, and use these to take a position in relation to a desire to carry out that action. However, the desire to carry out that course of action has equal claim to form part of a perspective that represents the authoritative self. The authoritative self is the holistic whole. Therefore, there is a continuity between the 'I' in each of its possible positions that is not implied by the kind of divided self involved in hierarchical accounts. There is not one, true, permanent position for the 'true' self to hold, but rather a plethora of positions spread across the entirety of the self. These are not defined in fundamentally different ways, and also can be absorbed into one another, as higher and higher perspectives are reached. This underlies a level of continuity that helps explain why different aspects of the soul can have traction upon one another, when it comes to the mechanism of self-governance. Reaching a higher level of autonomy involves incorporating a greater number of perspectives contained within the self into a higher perspective, or incorporating a greater range of relevant information contained

within the world into a judgement. This is continuous with the low autonomy state. Each of the kinds of materials that are present in the low autonomy state can be incorporated into the high autonomy state, and the self as a whole, not one particular and limited type of aspect, is relevant for securing the conditions of autonomy.

In a sense, Buss' (2013) concerns around integration that I discussed in chapter two, and reapplied to a critique of hierarchical accounts of the self-relation involved in personal autonomy, are simply not applicable, when it comes to the horizontal account. The different possible versions of the horizontal account do not operate on a model of self-alienation, by which the authoritative self distances itself from external possible motivational underpinnings of behaviour. This is a key feature of hierarchical accounts, that opens up the concerns discussed before around integration, since the authoritative self is able to become distanced from other possible motivational underpinnings of behaviour. On the other hand, when it comes to the 'horizontal' account, the entire self represents the authoritative self. Nothing is external to nor alienated from this single authoritative self, so there is no need to explain how independently operating contrary influences are able to be brought under the remit of the governing self. No aspects of the self are truly alienated from the possible purview of the governing self, in contrast to hierarchical accounts that fundamentally divide the self in a permanent, or semi-permanent way.

For Frankfurt's early (1971) account this self-alienation occurs when an agent forms a decisive higher-order volition that a specific desire, as opposed to a specific contrary desire, is effective in leading the agent to action. They are then alienated from the contrary desire, and will fail the conditions for autonomy if that is the one that is effective in leading them to action. Since the one desire (together with the corresponding higher-order volition) and the other by definition represent distinct concerns, for instance a concern with work and family commitments versus a concern with hedonic pleasure, it is hard to see why the latter should be swayed by the concerns of the former. Altogether, it is unclear why a decisive higher-order commitment to one set of influences should have the power to bring contrary motivational influences into line. The enforcement powers of the agent's decisive higher-order volitions over the rest of the self as a whole are unclear. When it comes to character accounts (e.g. Ekstrom 1993), or accounts that focus on an agent's set of core values or commitments (e.g. Watston 2004, Frankfurt 2002 2006) or rational nature (e.g. Raz 2002), the authoritative self is defined in a fundamentally different way to the rest of the self's possible motivational influences. Therefore, the kinds of motivational influences upon action

that lead an agent to successfully self-govern are of a different kind to those that operate upon an agent when they fail to act in an autonomous manner, or when they are alienated from the motivational underpinnings of their actions. Motivational influences that fall outside of an agent's character or set of core values will by definition involve distinct concerns to those contained within that character or set of core values. Similarly, motivational aspects of the self that operate independently of the self's understanding of the relevant reasons will by definition not be moved by a concern for reasons, the moving force of the authoritative reasons-responsive aspect of the self, as contained in accounts like Raz's (2002).

On the other hand, the 'horizontal' account does not suggest that failures of autonomy are caused by self-alienation, in such a way. Under the horizontal account, it is not to say that the governing self is distanced from other, alternative, motivational underpinnings of behaviour: it is rather that the self operates under one perspective, containing a given set of motivational influences, rather than other lower-autonomy perspectives, that may contain different motivational influences, or a more limited set of the same motivational influences. The very perspectives contained in low-autonomy states can form a part of high-autonomy states. For instance, say the low autonomy state involves an addict acting on the basis of a desire to take the drug, whilst overlooking or failing to absorb a large number of considerations against taking the drug. Their attenuated perspective is autonomy-undermining, when it comes to their behaviour. However, the kinds of materials contained within this perspective can also form part of a high-autonomy state. When they adopt a broad view in relation to the action-decision situation at hand, they may incorporate the desire to take the drug, and the various points that support the stance of the desire to take the drug (e.g. short-term pleasure, easing of pain of withdrawal symptoms), into a broad perspective that also contains many further considerations and incorporates more of their beliefs and commitments: predictions about the future given repeated instances of this kind of behaviour, commitments to their family and work, and so on. In this sense, the high-autonomy state has traction over the low-autonomy state because the low-autonomy state is actually absorbed into the high-autonomy state.

Alternatively, it might not be that the low-autonomy state is absorbed into the high-autonomy state – perhaps the high-autonomy state involves a wider perspective, but not in a manner that contains the low-autonomy state. However, if the self is therefore operating outside the influence of these contrary motivational influences, they do not offer a threat to effective self-governance. Under the first possibility, the self is not alienated from these influences – rather they are *incorporated* into the judgement. Under this second possibility, in which the contrary

motivational influences are overlooked, it is still not however the case that the self is alienated from them – it holds more of a passive relation to those influences, in a similar way to the way in which certain things must always be overlooked, wherever the self falls in the range from a low-autonomy to a high-autonomy state. Those influences could still fall under the perspective of the governing self, in other instances of action or in the case of other judgements that the self might take on.

The higher the autonomy of the individual, the more that is *contained* within the outlook of the governing self, whether that is matters that are for, or matters that are against, the final action decision, or judgement. This helps to explain how the mechanism of self-governance and integration can occur. High autonomy states can always contain features of alternative, low autonomy states, and there is no one set definition of a high autonomy state, that appeals to a limited set of aspects of the self. On the other hand, whichever level of autonomy an individual is operating under, they will always be overlooking certain things. However, if a given motivational influence is overlooked, it is clear why influences contained with the governing self, as opposed to the relevant contrary motivational influence, will be effective when it comes to autonomous action.

Conclusion

In conclusion, we should be moving away from the dominant type of account of autonomy, a 'hierarchical' account, towards a horizontal view of the self-relation involved in autonomy. In this chapter, I have put forth a set of related positive arguments in favour of the horizontal conception of the self, implied by the horizontal account of autonomy, that denies that the authoritative self can be limited to one distinct aspect, or set of aspects with a similar structure. Humans have a limited capacity for reflexivity, and as it is possible to theorise about the unscrutinised underpinnings of other people's judgements and behaviour, we must recognise that such unscrutinised underpinnings are always involved in our own judgements and behaviour. Therefore, no single and limited outlook can present itself as the final authoritative aspect of the 'true' self.

I have also outlined the ways in which the horizontal account is able to bypass the shortfalls of hierarchical accounts, as set out in chapter 2. The entire self is the true self, despite the ranges of levels of autonomy that can be associated with different perspectives contained within that true self. Therefore, nonautonomous action, or action that exhibits low levels of

autonomy, is a meaningful phenomenon, under the horizontal account. Moreover, the horizontal account does not bifurcate the self in the dramatic manner implied by hierarchical accounts, avoiding the issues hierarchical accounts faced in explaining how the authoritative self is able to gain traction over the self as whole. Altogether then, we should be moving away from hierarchical accounts, towards a horizontal account of autonomy. In the next chapter, I will summarise my findings across the previous three chapters, and explain the ways in which a horizontal account of autonomy is able to fulfil the requirements for an adequate account of personal autonomy that I set out in chapter one.

Introduction

In this thesis, I have argued that we should move away from the dominant type of view of personal autonomy – the 'hierarchical' view – towards a horizontal account of autonomy. I have based this on a set of criticisms of hierarchical accounts, together with arguments in favour of a horizontal conception of the self, including observations about the features of high-autonomy states versus low-autonomy states. I have set out a number of ways in which the detail of the horizontal account could be filled out. In this chapter, I will summarise my findings. I will then touch upon the ways in which in which a horizontal account of autonomy is able to satisfy the conditions for an adequate theory of autonomy that I set out in chapter one.

Summary of Findings

In chapter two, I provided a number of examples of the type of structure of account I refer to by the 'hierarchical' label. Such accounts identify a limited aspect of the self, or limited set of aspects of the self, that compose the core or 'true' self. Agents are autonomous when their actions or their judgements are in line with the dictates of these authoritative aspects of the self. Nonautonomous action occurs when contrary motivational impulses issue in behaviour.

However, these accounts fall prey to two key shortfalls: they are unable to adequately explain the possibility of nonautonomous action, or the mechanism by which the authoritative aspect of the self is able to gain traction over the self as a whole. They cannot adequately explain the possibility of nonautonomous action because they do not explain why the authoritative self would allow itself to be taken over by alien forces, or how, if this did indeed occur, the resulting behaviour could still be regarded as an action that could be ascribed to the agent as a whole. Moreover, they do not adequately explain the mechanism by which the relevant core aspects of the self are able to wield authority over the self as a whole, since the self-as-governed and the self-as-governor are defined in fundamentally different ways.

In chapter three, I outlined the possible shape of a 'horizontal' alternative. I detailed the way in which, under a horizontal account, the self can take on a number of opposing perspectives,

but it is not to say that any one has final claim to be the outlook of the authoritative self. Instead, the self as a whole is the authoritative self. It can take on different perspectives at different points in time, across the self as a whole, and also higher perspectives that can absorb and adjudicate between a number of lower perspectives. I also discussed the conditions for high-autonomy states, versus low-autonomy states, under a horizontal account of autonomy. I argued that this is to be tied to the richness and breadth of the perspective adopted by an agent, in relation to a given action-judgement, or other kind of stance or judgement that the self might take on. This helps explain the intuitive pull of reflexivity and 'all-things-considered' judgements when it comes to the conditions of autonomy, together with specific examples of autonomy-deprivation. Since no perspective is truly complete, it also means that the structure of high-autonomy states is not fundamentally different to the structure of low-autonomy states, and that autonomy is always a matter of degrees.

In chapter four, I supported the picture of the self implied by the horizontal account by appeal to the argument that the human capacity for reflexivity is always limited, and whichever stance we take, whether in relation to an action-decision or other kind of judgment, always involves the application of materials which are the *tools* of scrutiny, as opposed to the objects of scrutiny. These tools of scrutiny can always be criticised and assessed from other perspectives within the same self, undermining the notion that any particular aspect, or set of aspects of the self, can have final and authoritative claim to represent an agent's point of view.

I also discussed the ways in which the horizontal account is able to bypass the shortfalls of 'hierarchical' accounts that I discussed in chapter two. The horizontal account can explain the possibility of nonautonomous action through the agent's adoption of attenuated perspectives. These perspectives are not fundamentally different in kind to the kinds of perspectives involved in high-autonomy states (that will also be attenuated to some degree), and still involve elements that form part of the 'true self', and that, further, can also be adopted into high-autonomy states. The horizontal account rejects the model of alienation implied by hierarchical accounts, and autonomy is a matter of degrees. Moreover, the horizontal account can explain the kind of integration involved in the authority relation because the entire self is the authoritative self, and there is a continuity between low autonomy and high autonomy states. In particular, low autonomy states can be absorbed into high autonomy states.

Meeting the Conditions for an Adequate Theory of Autonomy

Finally, I will discuss the ways in which the horizontal account of autonomy is able to meet the conditions for an adequate theory of autonomy that I discussed in chapter one.

I stipulated that personal autonomy involves a capacity to lead a self-directed, or self-governed life, and that autonomous action issues from an agent in a deeper and a truer sense than nonautonomous action. Under the horizontal account of autonomy, autonomous action is to be explained by conditions that pertain to the state of the self as a whole, and moreover by the breadth of the perspective adopted by an agent, whether that is measured in relation to aspects of the self, or aspects of the world. The relevance of the self as a whole helps to support the contention that the horizontal account of autonomy can be used to explain the possibility of self-directed or self-governed action. Moreover, if the breadth of the agent's perspective in relation to a given action judgement is to be measured in relation to features of the self, there is a clear and intuitive explanation of why more autonomous action is more truly the agent's own action. When the relevant benchmark is the world, it might seem less clear why more autonomous action is to be more truly considered the agent's own. However, the explanation fits intuitive examples of autonomy-deprivation, such as the addiction example and examples in which individuals have a limited set of options to choose from, and emphasis should be placed upon the fact that the self is entirely a product of the world. Therefore, the relation between the self and the world can have valid claim to be determinant of the extent to which an action is self-determined.

I also stipulated that an adequate account of autonomy should explain how nonautonomous action can be possible. As discussed above, the horizontal account can achieve this by explaining low-autonomy states in a manner that is not fundamentally different to the explanation of high-autonomy states, and by allowing each possible motivational underpinning of action to have valid claim to form part of the 'true', authoritative self. I mentioned that an adequate account should explain the mechanism by which successful self-governance can be wielded. For the horizontal account, this involves taking a wider and richer perspective in relation to a given judgement or action-decision, but a perspective that can also absorb elements contained in narrower, lower-autonomy perspectives that might also be adopted by an agent.

I also mentioned that an adequate account should be able to explain the kinds of autonomy-deprivation involved in addiction, inadequate rational and reflexive capacities, and a limited

framework of choice. These can all be explained by the kinds of attenuated perspectives involved in low-autonomy states, under the horizontal account. For instance, inadequate rational and reflexive capacities mean that an individual is not able to conceptualise the same kinds of abstract possibilities and goods that are involved in the perspectives of the highest autonomy states.

Conclusion

Altogether, then, we should be moving away from a 'hierarchical' account, towards a horizontal account of personal autonomy. The precise details of such an account are beyond the scope of this thesis. However, the horizontal account argues that self-governance involves relations between aspects of the self that are fundamentally equal, and that autonomy is to be tied to the richness and the breadth of the perspective adopted by an agent, in relation to a given judgement or action decision.

6. References

Nomi Arpaly (2000): 'On Acting Rationally against One's Best Judgment', *Ethics* Vol. 110, No. 3 (April 2000), pp. 488-513

M. Bakhtin. (1973): *Problems of Dostoevsky's Poetics* (2nd ed.; R.W. Rotsel,Trans.). Ann Arbor, MI: Ardis. (Original work published 1929)

Sarah Buss (1994): 'Autonomy Reconsidered', *Midwest Studies in Philosophy*, XIX

Sarah Buss (2013): 'The Possibility of Action as the Impossibility of Certain Forms of Self-Alienation', book chapter, *Oxford Studies in Agency and Responsibility Volume 1*, edited David Shoemaker

Sarah Buss and Andrea Westlund (2018): "Personal Autonomy", The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/spr2018/entries/personal-autonomy/

G, Dworkin (1970):, "Acting Freely," Noûs, 4: 367–83

Laura Ekstrom (1993): 'A Coherence Theory of Autonomy', Philosophy and Phenomenological Research Vol. 53, No. 3 (Sep., 1993), pp. 599-616

Fichte (SE): *Das System der Sittenlehre nach den Principien der Wissenschaftslehre,* trans. Daniel Breazeale and Günter Zöller. *The System of Ethics in accordance with the Principles of the Wissenschaftslehre*. Cambridge: Cambridge University Press

Harry Frankfurt (1971): 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, Vol. 68, No.1 (Jan. 15 1971)

Harry Frankfurt (2002): Reply to Moran, from *Contours of Agency: Essays on Themes from Harry Frankfurt*, ed. Sarah Buss and Lee Overton, MIT Press

Harry Frankfurt (2006): 'The Question: "How Should We Live?" from *Reasons of Love*, Princeton University Press

Heidegger: SZ – *Sein und Zeit*, Tu¨bingen: Max Niemeyer Verlag, 1993.

Hermans, H. J., Kempen, H. J., & Van Loon, R. J. (1992). 'The dialogical self: Beyond individualism and rationalism.' *American Psychologist*, 47(1), 23–33. https://doi.org/10.1037/0003-066X.47.1.23

Hubert J M Hermans (2001): 'The Dialogical Self: Toward a Theory of Personal and Cultural Positioning', *Culture and Psychology* 2001; 7; 243

Gene M. Heyman (2010): *Addiction: A Disorder of Choice*, Harvard University Press

Higgins, S. T., Delaney, D. D., Budney, A. J., & Bickel, W. K. (1991): 'A behavioral approach to achieving initial cocaine abstinence.' *American Journal of Psychiatry, 148*, 1218–1224

Higgins, S. T., Wong, C. J., Badger, G. J., Ogden, D. E., & Dantona, R. L. (2000): 'Contingent reinforcement increases cocaine abstinence during outpatient treatment and 1 year of follow-up.' *Journal of Consulting and Clinical Psychology, 68,* 64–72.

John Hyman (2015): *Action, Knowledge, and Will*, OUP

W. James (1890): *The Principles of Psychology* (Vol. 1). New York: Henry Holt.

Wallace, R. Jay (2001). "Normativity, Commitment, and Instrumental Reason." *Philosophers' Imprint 1* (3): 1–26.

CM Korsgaard (1996): *The Sources of Normativity*, CUP

Jonathan Lear (2008): *Radical Hope: Ethics in the Face of Cultural Devastation*, Harvard University Press

Denis McManus (2018): 'On a Judgment of One's Own: Heideggerian Authenticity, Standpoints, and All Things Considered', *Mind*, Vol. 128 . 512 . October 2019

Richard Moran (2001): *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton University Press

Peter Railton (2012): 'That Obscure Object, Desire', *Proceedings and Addresses of the American Philosophical Association 86* (2):22-46 (2012)

Joseph Raz (2002): 'When We Are Ourselves: The Active and the Passive', chapter 1 from *Engaging Reason: On the Theory of Value and Action*, OUP

Waldorf, D., Reinarman, C., & Murphy, S. (1991). *Cocaine changes: The experience of using and quitting*. Philadelphia: Temple University Press.

Owen Ware (2015): 'Agency and Evil in Fichte's Ethics', *Philosophers' Imprint 15* (2015)

Gary Watson (2004): 'Free Agency' from *Agency and Answerability: Selected Essays*, OUP

Bernard Williams (1979): 'Internal and External Reasons', in Ross Harrison (ed.), *Rational Action.* Cambridge University Press. pp. 101-113

Richard Yetter Chappell & Helen Yetter-Chappell (2016): 'Virtue and Salience', *Australasian Journal of Philosophy* 94 (3):449-463