The concept of autonomy

by Ian Jennings

submitted towards the degree of Master of Arts in Philosophy at the University of Cape Town in February 1996

****

Introduction

The question of which of our actions or desires are genuinely attributable to us is the question I examine in this thesis. I use the term "autonomous" to describe those agents whose desires or actions are genuinely their own, and I refer to actions or desires which cannot genuinely be attributed to agents as heteronomous actions or desires.

I have chosen to discuss this question under the rubric of the concept of autonomy, although the number of near-synonyms in the philosophical literature means that I could, perhaps, have referred instead in my title to concepts such as freedom, responsibility, independence, authenticity, self-determination, self-identity, freedom of the will and similar concepts. But whatever terminological choice is made, the issue that interests me concerns the nature of those actions or desires which are genuinely the agent's - those desires and actions which, as some have put it, are the agent's *real*[1] desires and actions.

The concept of autonomy is crucially important for both moral and political ideals. In his paper `Autonomy and personal history'[2] John Christman says that

> [virtually any appraisal of a person's welfare, integrity, or moral status, as well as the moral and

---

[1] Susan Wolf, for example, uses this term in *Freedom within reason*, but her use of the term is a more restricted one, in that she refers to a particular theory of autonomous desires as the "Real self" theory.

[2] (1991) 21 *Canadian journal of philosophy* 1.

political theories built on such appraisals, will rely crucially on the presumption that her preferences and values are in some important sense her own. In particular, the nature and value of political freedom is intimately connected with the presupposition that actions one is left free to do flow from desires and values that are truly an expression of the "self-government" of the agent.

Our concern with political injustice is typically concerned not only with the physical and mental suffering imposed on its victims. The claims of racists or sexists that blacks or women are happier in their subservient roles are not generally taken as sufficient to justify the oppressive practices they are intended to justify. This is because we believe that the autonomy of blacks and women is violated by political oppression. And it is from this concern with autonomy that our notion of human rights springs.

Our concerns with guilt and justice are likewise dependent on the notion of autonomy. We hold typically that, for a person to be found guilty in a court of law, they must be shown to have been acting autonomously. This belief is usually expressed in the view that one cannot be held morally responsible for actions which one "could not help", as such actions cannot genuinely be attributed to one. Conversely, a guilty person is one whose actions can genuinely be attributed to them

My plan in the main body of the thesis is to discuss, in chapter one, what I call the existentialist view of autonomy: the view that the mark of autonomous agents is that they have the capacity to exercise ultimate control over their actions and desires. I argue that this view is an inadequate conception of autonomy because of the incoherence and the

implausibility of the conception of freedom as the absence of causal determination.

In the second and third chapters I discuss what I have termed subjective views of autonomy. There are a number of different subjective views - I discuss in particular the theories of Harry Frankfurt, Gary Watson and J David Velleman. What these theories have in common is that they do not regard the capacity to exercise ultimate control as the mark of autonomy, but instead see autonomous action as issuing from agents who *identify* with their motivations in ways which each writer specifies. What is typical of subjective views is that they take the agent's appraisal of their own desires as crucial to an assessment of their autonomy. And it is this feature, I argue, that gives rise to certain serious problems for subjective views, and which leads me into the discussion of the views of Susan Wolf and John Christman in the final chapter.

Both Wolf's theory and Christman's theory accept the premise that ultimate control is not what confers autonomy, and both are efforts to overcome the difficulties with the subjective views which I highlight in the second and third chapters. However, they approach these difficulties in notably different ways. Wolf argues that it is the capacity to act in accordance with reason that confers the status of autonomy, in my terms, while Christman argues that the process of desire-formation must fulfil certain conditions in order for the person whose desires they are to be considered to be autonomous. This approach, I will argue, is the one most likely to lead to a coherent and plausible conception of autonomy.

****

Chapter 1 : The existentialist view of autonomy

A successful philosophical account of autonomy cannot be content with listing the characteristics of an autonomous person. It needs to explain *why* such characteristics confer autonomous status on the person who displays them, and in doing so must also explain why certain characteristics are marks of loss of autonomy. In this chapter I wish to examine one attempt at providing such an account. I will refer to this particular attempt as the *existentialist view of autonomy* because of its apparent connections with Sartre's picture of freedom,[3] although the features of this view which I am concerned with - in particular its emphasis on the requirement that one be able to do other than what one in fact does, all other circumstances remaining the same - are often referred to in the literature on free will as the *libertarian* view.[4] I do not intend my use of the label "existentialist view" to indicate a scholarly interest in the details of Sartre's views and nothing I say in this chapter is meant as an attempt to contribute to the interpretation of the existentialist tradition.

Roughly speaking, the existentialist view suggests that in order for us to be autonomous we need to have *ultimate control* over our actions, in the sense that no force which originates outside ourselves compels us to choose one course of action rather

---

[3] *Elbow room* 83.

[4] Double *The non-reality of free will* - especially chapter eight.

than another. Agents with ultimate control determine the content of their wills, and they themselves are not determined by anything external to themselves. Susan Wolf refers to this ability as "autonomy" and the view that ultimate control is necessary for autonomy as *the autonomy view*[5], but I will not be using her terminology[6] given that in this thesis I use the word autonomy in a broader sense. In what follows I will present considerations in favour of the existentialist view and assess whether they stand up to scrutiny.

Characteristics of autonomous agency

The existentialist view is intuitively plausible - hence my decision to discuss it before any of the alternative views of autonomy. Let us examine what might lead us to adopt it:

I begin by following Susan Wolf's procedure of examining various characteristics of people whom we would naturally regard as autonomous, on the grounds that it seems likely that we could then extrapolate a general principle from the sample. I should say at this point that Wolf uses the word "responsible" where I use the word "autonomous" , but since she means by "responsible"[7] what I mean by "autonomous", in what follows I shall

---

[5] *Freedom within reason* 10.

[6] Unless explicitly indicated otherwise.

6

Responsibility, according to Wolf, means 'enjoying a relation to one's will which makes it the case that one's actions are up to oneself'. Cf *Freedom within reason* 4.

use the vocabulary of autonomy and responsibility interchangeably.

The most obvious characteristic of autonomous agents is that they possess a will which is the source of their actions. We do not credit inanimate objects with autonomy, and we do not regard events which lie beyond an agent's power to control as the autonomous actions of that agent. Wolf describes this feature of autonomous agents as the possession of a potentially effective will. In her terms, one can be held responsible only for '...events and properties that stand in a relation to [one] such that [one's] will is or could have been *effective* in determining the existence of these events or properties.'[8] A person could not, for example, be held responsible for the occurrence of an earthquake, given that their desire for an earthquake to occur could not be what brings it about. On the other hand, they could be held responsible for failing to warn others of an impending earthquake if they were aware of this possibility. Warning others clearly falls within the compass of the potentially effective will.

The second characteristic is equally uncontroversial, although perhaps a little less obvious. The need for it becomes clear when one realizes that there are many creatures with potentially effective wills which we do *not* regard as autonomous agents. Animals, for example, and small children, fall into this category. Such creatures do have control of a certain kind over their behaviour, in that they experience desires and are capable of acting to satisfy those desires. But they are unable to control their behaviour, in Wolf's terms '...with respect to the features that would be relevant to the judgements of

---

[8] *Ibid* 7.

responsibility.'[9]

An example of an action which we wouldn't hold someone fully responsible for, but which was nevertheless the product of a potentially effective will, would be that of a golfing beginner who scores a hole-in-one. We wouldn't say that what the beginner did was uncontrolled behaviour, because he or she had the desire to hit the ball into the hole and was in fact able to act in a way that satisfied that desire. We recognize, though, that the beginner could not have been influenced by the relevant considerations, and so we attribute the hole-in-one to luck. We don't hold the beginner responsible in the way that we would hold Nick Faldo responsible for a hole-in-one. This second characteristic of actions which are due to the agent - that someone be able to control their behaviour in an intelligent way - Wolf calls the condition of relevant intelligence.[10]

The conditions of a potentially effective and a relevantly intelligent will are, however, still not sufficient for autonomous agency. This is shown by the fact that it is possible to think of examples of agents who act possessing both effective and relevantly intelligent wills, but whose actions are nevertheless not generally taken to be autonomous.

An example of this would be the case of someone who is coerced to act in a certain way because their life is threatened. Imagine, for example, that confidential information is demanded of someone at gunpoint. Most of us, if we found ourselves in this situation, would probably disclose the information demanded. In doing so we might be acting

---

[9] *Ibid.*

[10] *Ibid* 8.

deliberately, understanding exactly what it is we are doing, thus demonstrating an effective and relevantly intelligent will, and yet, again, our actions would not be taken to be those of an autonomous agent.[11] A kleptomaniac would be another example of an agent who lacks autonomy despite possessing an effective and a relevantly intelligent will. People who suffer from this disorder act intentionally, knowing that what they do is wrong. They are not, however, held to be acting autonomously when they steal.[12] They are understood to be acting under influences which they cannot control.

In all of these cases the problem appears to lie in the *source* of the agent's will. Although their behaviour is controlled by their wills, their wills are not controlled by them. They therefore appear to lack ultimate control over their behaviour. And this, it might seem, serves to deny them the status of autonomous agents. In Isaiah Berlin's words, if it is *not* the agent's control which confers autonomy, then

> ...what reasons can you, in principle, adduce for attributing responsibility or applying moral rules to [people] which you would not think it reasonable to apply in the case of compulsive choosers - kleptomaniacs, dipsomaniacs, and the like?[13]

In other words, if the difference between responsible/autonomous people and compulsives, such as kleptomaniacs and dipsomaniacs, is not one of control, then it's

---

[11] *Ibid* 9.

[12] *Ibid.*

[13] *Four essays on liberty* 20-1.

hard to know what the difference might be. And it seems, at first blush, that to have control over our wills, we must be free from outside influences.

Ultimate control

In the light of this, many have felt that it is natural to suppose that there is a third characteristic necessary for autonomous agency in addition to the possession of a potentially effective and relevantly intelligent will. The third characteristic would be the condition that one be in ultimate control of the desires one acts on. To have ultimate control would be to be free from outside influences. So agents with ultimate control would act in the light of their desires and choices, and these desires and choices would arise from within themselves. Nothing external to themselves would *make* them choose anything - they would be able to follow a course of action or not follow it without being determined to follow either option by outside forces. This fits with the common sense idea that to be considered an autonomous being one must have a kind of control over one's behaviour which typically heteronomous beings such as lower animals and children don't have. And it fits with the feeling we have that we are less responsible for those actions we are compelled in some way to perform than those which we choose to perform.

A number of writers, however, have questioned whether ultimate control is the

feature of an agent's make-up that confers autonomy upon him or her, suggesting that there are enough problems with the existentialist view to force us to look elsewhere for an account of autonomy. In what follows I want to look at the arguments against the existentialist view under three broad headings.

Firstly, I wish to examine the view that the capacity for ultimate control is a capacity we have no reason to desire. Secondly, I will move on to discuss the view that the capacity for ultimate control is one which is unnecessary for autonomy. And thirdly, I will come to the view that ultimate control is in fact impossible. Finally, I will assess what the implications of these criticisms are for the existentialist view in general.

****

Is ultimate control something we have reason to desire?

The view that ultimate control is an undesirable capacity is one that many philosophers have held. I will concentrate here on the reasons Susan Wolf[14] and Daniel Dennett put forward for holding this view.

Susan Wolf argues that having the capacity to exercise ultimate control over one's

---

[14] In *Freedom within reason*, in particular chapters three and four.

choices means having the ability to choose contrary to reason. This, she says, is an ability that we have no reason to desire.

Her argument goes as follows: If one has ultimate control over one's choices, one is not constrained by any external forces to choose in any particular way. One chooses just as one pleases. This ability, she says, means that one is not bound to make any choice which appears more reasonable than its alternatives, simply because, as an agent who possesses ultimate control, one is not bound by anything at all in making one's choices. Ultimate control of one's choices goes beyond the mere ability to act in accordance with reason. One can make what Wolf calls 'radical choices'.[15] Such choices can be made for no reasons, or even for bad reasons, when one is aware that there are in fact good reasons for making a particular choice. As she puts it:

> Since a radical choice must be made on no basis and involves the exercise of no faculty, there can be no explanation of why or how the agent chooses to make the radical choices she does.[16]

It is important to note that desiring the ability to make radical choices should not be confused with desiring the ability to respond differently to different situations. Wanting the latter ability is wanting to be able to respond rationally to whatever different circumstances one encounters. It is, in other words, wanting to avoid irrational or excessive rigidity in

---

[15] *Ibid* 53.

[16] *Ibid* 54.

one's responses to the world. Wanting the ability to make radical choices, on the other hand, is wanting the ability to respond rationally or irrationally to one's circumstances, just as one so wishes.

In summary, then, the aim of Wolf's argument is to show that we do not have any reason to want the capacity to exercise ultimate control. To reinforce this conclusion she describes a case in which there is an obviously rational course of action available and attempts to argue that in such a situation we would not have reason to want an ability to act in any other than the rational way.

In her example she describes two people on a beach, one of whom has ultimate control over their actions, and another who is unable to act contrary to reason.[17] This is the only difference between the two. The situation they find themselves in is one where they both hear the cries of a child who is in difficulty in the water. Both of them (simultaneously) attempt to save the child, both believing that this is the only rational thing to do in the circumstances. The only difference between them is that the person who has ultimate control is capable of acting in defiance of the recognition that right reason obliges them to try to rescue the child. Her question to those who adhere to the existentialist view of autonomy is why anyone would want to be the agent with ultimate control in this situation.

---

17 *Ibid* 58 *et seq.*

Dennett raises the same question :

> Suppose I am in the supermarket, trying to decide which can of soup to buy. I have heard disturbing rumours about the tricks advertisers use to "control" my buying habits, and I certainly don't like that idea. But do I then hope that when I get to the soup shelves *nothing* will control my purchase decision? Do I want the sort of "radical freedom" that would make me impervious to important and relevant features of the candidates?...[S]houldn't I be content to let my choice be "controlled" by quality, price and availability...?[18]

On the basis of these examples it is hard to understand why anyone would want ultimate control. But are we being misled in some way? Are there reasons why we might still want this ability?

Why one might want to have the ability to be irrational

There are a number of suggestions as to why there could be reason for wanting the ability to act contrary to reason. In the following section I will discuss a number of these suggestions.

I) Reason as a stereotyped way of thinking

---

18 *Elbow room* 65-6.

Firstly, one may associate reason with a particularly narrow style of thinking, or perhaps some culturally or sexually oppressive standard whereby creativity and sensitivity are disregarded. If one holds to a view like this one would naturally wish to retain the ability to act contrary to reason, thus understood, as a defence against a whole number of evils, intellectual and otherwise.

But such understandings of reason are narrowly polemical. They are not all *per se* attacks on the idea that one ought to act for good rather than bad reasons, and, if they are such attacks, it's hard to see how they could convince anyone without appealing to reasons. The issue of the nature of reason is a highly discussed one in contemporary thought - particularly in the philosophy of science.[19] For present purposes I will simply suggest, however, that adopting a broader, normative definition of reason - for example, as the optimal thinking process for reaching true beliefs about the world[20] - causes this objection to fall away.

ii) Following reason is a restriction of one's liberty

Another objection to the view that we have no reason to desire ultimate control is the claim that one's freedom would be restricted if one *had* to follow reason. Clearly, we

---

[19]See, for example, the work of Paul Feyerabend.

[20] Cf *Freedom within reason*, page 56, where she says:

> "Reason", as it is being used in this book...is an explicitly and essentially normative term. It refers to the highest faculty, or set of faculties, there are - that is, to whatever faculties are properly thought to be most likely to lead to true beliefs and good values.

desire freedom, and so perhaps this is why we have reason to desire ultimate control. Reason obviously precludes certain choices as being irrational, and it may be that this ought to be seen as an unwelcome restriction on our freedom. Wolf puts this view as follows :

> ...[I]f, being rational agents, it is not up to us to have and to act on the reasons we have, then action in accordance with Reason is no more autonomous than action in accordance with any other psychological process would be.[21]

Why might one accept such a claim? On the face of it, many of our experiences of choosing - experiences in which we often consider ourselves to have control over our choices - do suggest that in fact we are restricted in our options by reason. In such instances we generally regard ourselves as making the choices we do in the light of the facts available to us. Given certain facts, we make certain choices. And this leads naturally to the supposition that, once particular facts become obvious to us, we cannot but make the choices we do if we are to be rational. So, one might conclude, rationality is an undesirable restriction on our freedom of choice.

And one might therefore want the ability to act contrary to reason because, amongst the many things one regards as valuable (including, presumably, reason), one also values one's freedom of choice.[22] As Wolf points out, though, there seems

---

[21] *Ibid* 52.

[22] *Ibid* 57 *et seq.*

something odd in wanting an ability one has no intention of ever using. Arguing that valuing one's freedom requires that one has options one would never use seems analogous to feeling restricted because the police will not allow you to become an inmate of a prison should you so wish.

The reason we think being free is more desirable than being dictated to is because we don't want to be pushed in directions which would be unpleasant or bad for us. It seems obvious that our best defence against this is reason itself. Although we might, at the time of a conflict between reason and desire, feel that being compelled to be rational was undesirable; our reflective and considered desire would undoubtedly be that the ability to follow our wanton and self-destructive urges was something we would rather be rid of, and that being compelled to be rational was not, in the final analysis, undesirable at all. As David Wiggins puts it:

> The libertarian ought...to be content to allow the world, if it will only do so, to dictate to the free man how the world *is*. Freedom does not consist in the exercise of the...right to go mad without interference or distraction by fact.[23]

Similarly, Dennett asks rhetorically if we knew that we were being controlled by a benevolent counterpart to Descartes' evil demon - a friendly all-powerful advisor - 'who used only epistemically warranted communicative interactions to achieve only cognitive

---

[23] *Towards a reasonable libertarianism* 34.

effects'[24], whether we would have any reason to find this condition undesirable. It's hard to see why this could be understood as coercion, since I am only coerced if I am prevented from taking into account some relevant consideration.[25]

****

Is ultimate control necessary for autonomy?

If ultimate control is something we have no reason to desire, it may yet be necessary for autonomy. After all, one might well find freedom undesirable. I want to argue, however, that not only is ultimate control undesirable, it is also unnecessary for autonomy. In making the same point, Susan Wolf uses the example of the two swimmers who enter the water to save the drowning child to argue that having or lacking ultimate control does not affect the moral status of agents. Both swimmers are equally morally praiseworthy, she argues, in that both do the right thing for the right reasons, and the fact that one of them lacks ultimate control does not mean that she is any less praiseworthy than the other one. Given that attributing moral praise and blame are appropriate only to

---

[24] *Elbow room* 65.

[25] Berofsky 203-4.

autonomous agents, she concludes that there is no difference in autonomy between the two.

One possible objection to the suggestion that ultimate control is unnecessary for autonomy is that the person with ultimate control displays a greater degree of autonomy in the overcoming of temptation, and is therefore to be accorded a deeper moral status. We generally do feel that praise is more appropriate towards a person who does the right thing in circumstances which make it difficult than towards someone who does so in easy circumstances.

But the two people under discussion can't be said to be in different circumstances. As the example was set up, the difference between the two lies solely in the first agent's ability to reject reason. Both people, then, could have had a battle with temptation - or both could have found acting in accordance with reason easy. The only difference is that the person who lacks ultimate control could not have chosen to act irrationally whereas the person who has ultimate control could.

Another, perhaps more plausible, suggestion regarding the superior autonomy of agents with ultimate control is that agents who lack it act mindlessly, or mechanically, and cannot therefore be autonomous. As Joel Feinberg puts it 'Those who most conspicuously fall short of...autonomy are not those who are wicked, but rather those whose "morality" is a mindless reflex.'[26] This suggestion appears to imply that someone who lacks ultimate control cannot act rationally, as acting "mechanically" is usually

---

[26] *The moral limits of the criminal law* 39.

contrasted with acting rationally.

But whatever one takes the term "mechanically" to mean, there is no reason to suppose that a person who lacks ultimate control can't follow subtle and sophisticated methods of reasoning. Wolf's example assumes only that they couldn't help doing so. And if they did follow such methods of reasoning, it is hard to understand why their behaviour should not be regarded as autonomous.

However, one may still object that making the right choice for the right reasons doesn't confer any responsibility or dignity on the agent unless that choice is *truly theirs*. This is a legitimate concern, but one which, in the present context, is voiced misguidedly. Believing that one needs the ability to act contrary to reason before any reasonable choices one makes can truly be said to be one's own, implies that if one is determined by reason to make a particular choice, then that choice cannot be said to be one's own.[27]

And if this view is correct, it's hard to imagine which choices could be truly one's own. Certainly not choices which are made on no basis whatsoever, as there would be no sense in which such choices were controlled by the person concerned. Perhaps one could believe that choosing on the basis of one's desires confers the requisite responsibility or dignity on the choice made. But one would have to be certain that the desires were not subject to any external influences, and we will later come to see that this is not a very likely prospect.

It therefore seems, says Wolf, that the fact that the person in her example with

---

[27] Wolf *Freedom within reason* 58.

ultimate control can reject her reason in addition to her ability to act in accordance with it adds nothing of value to her condition in virtue of which her actions can be seen to take on a greater significance than the actions of the person who was unable to reject her reason.[28]

These considerations are further reinforced by Harry Frankfurt's claim that the existentialist view fails to account for our common sense view of autonomous agency. He points out that such a view fails to explain why we are unwilling to allow that members of species inferior to our own have what he calls freedom of the will[29] and what I have called autonomy.

If one views autonomy as the absence of external determination, then one must understand each instance of autonomous action as miraculous - an event uncaused by physical events. If this is true, it is not clear why this is an ability only those whom we normally take to be autonomous agents enjoy. Creatures which we would not regard as autonomous appear to initiate actions in much the same way as those we do so regard. On this view, a dog's action in lifting its paw is as miraculous as any human action.

In summary, Wolf makes it clear that there is no reason to suppose that someone who couldn't help being rational could not be a responsible/autonomous agent. She has

---

[28] Wolf *Freedom within reason* 62.

[29] Frankfurt 23.

shown, therefore, that ultimate control therefore cannot be a necessary condition for autonomy. And, as I will argue below, ultimate control cannot in fact confer autonomy at all.

****

Can ultimate control confer autonomy?

It seems to me that the strongest case against the existentialist view of autonomy lies with the traditional compatibilist objections to the view that freedom requires the absence of causal determinism. I will present these objections below.

First of all, it seems quite unlikely that we do in fact possess ultimate control. This becomes apparent when we ask ourselves what the source of our desires and actions is. Initially we are able to postulate that the source lies within us, but we are always able to ask why that particular source arose, and we are usually able to find an answer. In the end we usually concede that we are to some extent products of our heredity and our environment.

But if all of our actions are the result of forces which can ultimately be traced to sources outside of ourselves in our heredity and our environment, then it is hard to see

how any of our actions could be ultimately controlled by ourselves. And if ultimate control is indeed a condition for autonomy, it is hard to see how anyone could ever be autonomous. This is a troubling possibility. But it is not in itself a refutation of the existentialist view - it may simply be the truth that we are not autonomous beings.

The more telling objection is this. If we try to imagine behaviour which meets the condition of ultimate control we are presented with a puzzling picture. It is not at all clear that the possession of ultimate control, even if it were possible, would in fact confer autonomous status on us.

As Richard Double points out[30], what I term existentialist theories have difficulty in explaining how agents who have the ability to choose either way, all antecedent conditions remaining the same, can regard the outcomes of their choices either as rational, or under their control. To put it in his words:

> If the agent might either make a choice or do otherwise, given all the same past circumstances, and the past circumstances include the entire psychological history of the agent, it would seem that no explanation in terms of the agent's psychological history, including prior character, motives and deliberation, could account for the actual occurrence of one outcome *rather than* the other, ie, for the choosing rather than doing otherwise, or vice versa...[T]he outcome may be different...though the psychological history is the same.
>
> ...[W]hat I cannot understand is how I could have reasonably chosen to do otherwise,

---

[30] *The non-reality of free will* chapter eight.

> how I could have reasonably chosen B, *given exactly the same prior deliberation* that led me to choose A, the same information deployed, the same consequences considered, the same assessments made, and so on.[31]

It seems reasonable to assume that what makes an action rational is its connection with a particular set of deliberations? But if that same set of deliberations could give rise to two different actions, or two different thoughts, then it's hard to know what the connection between the deliberations and the actions are. The outcomes of the thought processes of agents with ultimate control begin to look just like chance happenings. Chance happenings are particularly unpromising candidates for the marks of autonomy, as the concept of autonomy implies that we can in some way hold autonomous agents *responsible* for their actions. But if these actions are chance happenings, then the agent cannot be held responsible for them.

The same problem arises in the case of the notion of control. If we are to hold agents responsible for their actions, we need to know that they controlled the actions concerned. This is simply the condition of an effective will that I mentioned at the beginning of the chapter. Bit if the same set of deliberations could give rise to two different actions, it's hard to know how I can be said to have controlled either, as there appears to be no connection between the deliberations and the outcomes.

So it appears that the possessing the ability to choose either way, all antecedent

---

[31] *Ibid* 194.

conditions remaining the same, an ability they would possess if they had ultimate control, then their actions could be neither rational nor under their control. And these are two considerations which we generally regard as being necessary conditions of autonomy.

But if it is true that ultimate control is necessary for my behaviour to be genuinely attributable to me then it would seem that autonomous behaviour could issue only from a self whose will was not influenced by external forces. But such a self would necessarily have no origin. Any attempt to explain what the origin of such an autonomous self is would necessarily fail, because to explain why it is what it is would attribute its nature to external forces. To be a truly autonomous self it would have to have arisen out of nothing.

The autonomous self would then look like either a random occurrence or an inexplicable manifestation. And it is difficult to see what the virtue in having one's will controlled by such a self would be. There is no reason to suppose that the actions of such a self are due to it in any deeper sense than the actions of a self which is the product of external forces, as it could hardly be said that anyone had *chosen* a self of this kind. It appears to be no more likely to confer the status of autonomy than an externally produced self.[32]

Ultimate control appears to be either impossible for agents to achieve, as it would be if our selves are always determined from outside themselves, or, worse still, to be an incoherent concept, as appears to be the case if our selves are arbitrary and accidental.

---

[32] Frankfurt 13.

In these circumstances it seems prudent to look elsewhere for a plausible account of autonomy.

Harry Frankfurt claims to offer us an account of autonomy which is not wedded to any assumptions about freedom and determinism because it leaves open the question as to how people come to enjoy freedom of the will. His account, he argues,[33] is compatible with the supposition that certain people enjoy freedom of the will as a result of a chain of natural causes or that their freedom of the will has come about by chance - or even in some other way, if there is some other way. It is to his views that I turn in the following chapters.

****

[33] *Ibid* 25.

****

Chapter 2 : Subjective views of autonomy : The early Frankfurt and Watson

It appears, as a result of the discussion in the previous chapter, that the existentialist view of autonomy - the view that autonomy consists in having ultimate control over one's actions - is untenable. The difference between actions which we would regard as autonomous and those we would not cannot be that the latter result from influences which fall outside the ultimate control of the agent and the former are not. It is not necessary, however, to conclude from this that the concept of autonomy is an incoherent one, provided we can come up with an alternative explanation of the difference between autonomous and heteronomous action.

Subjective views of autonomy

One type of response to this challenge is the collection of views I will term *subjective views* of autonomy. Views of this kind have in common the fact that they, to use John

Christman's phrase, include conditions of self-appraisal.[34] In other words, subjective views regard agents' own assessments of their behaviour as crucial in deciding whether they are autonomous or not. This is in contrast to both the existentialist view and those views in which the autonomy of behaviour is decided by facts in principle accessible to observers, regardless of the assessments of the agents.

My aim in the following two chapters is to examine a number of subjective views, assessing whether any version succeeds as an account of autonomy where the existentialist account fails. In doing so I wish to focus principally on the work of three writers: Harry Frankfurt, Gary Watson and J David Velleman. All three are concerned in some way with the problem of when actions can truly be attributed to agents, and all endorse some form of the subjective view. Not surprisingly, though, these different proponents of subjective views take different kinds of self-appraisal to be decisive in conferring autonomy on agents.

For Watson, autonomous action is informed by desires which conform to the agent's evaluations, while for Frankfurt, autonomous action is motivated by desires which are endorsed by the agent, either by means of another desire, as in the earlier papers, or by a decision, as in the later. Velleman, on the other hand, takes autonomous action to be action which conforms to the desire to act according to reasons.

In what follows I examine each view in detail, looking initially at Frankfurt's early

---

[34] (1991) 2 *Canadian journal of philosophy* 18.

papers 'Freedom of the will and the concept of a person'[35] and 'Identification and externality,'[36] in which he proposes what has come to be known as an hierarchical account of autonomous action. I then examine Watson's response to this paper, which shares certain characteristics with hierarchical accounts and then move on, in the following chapter, to a discussion of Frankfurt's later refinement of his position, as outlined in 'Identification and wholeheartedness,'[37] where he ceases to accord hierarchies of desire the importance he did in his earlier work. Following this I look at Velleman's paper 'What happens when someone acts?' and at the end of chapter four I assess the success of subjective views in explaining what the difference between having and lacking autonomy is.

Harry Frankfurt: Freedom of the will

Harry Frankfurt has offered a particularly sophisticated version of the subjective view in 'Freedom of the will and the concept of a person,' which serves as a good starting point. In this article, he proposes the view that I enjoy freedom of the will when I act on those of my desires by which I desire to be moved. His interest in 'freedom of the will' is

---

[35] Chapter 2 of *The importance of what we care about.*

[36] *Ibid* Chapter 5.

[37] *Ibid* Chapter 12.

identical to my interest in the phenomenon which I have called 'autonomy.' There may, of course, be other uses of the terms on which they are not synonymous, but both Frankfurt and I use them for the same purpose, namely to describe agents whose actions can be fully attributed to them. The following comments of Frankfurt's bear this out:

> The enjoyment of a free will means the satisfaction of certain desires...whereas its absence means their frustration. The satisfactions at stake are those which accrue to a person of whom it may be said that his will is his own. The corresponding frustrations are those suffered by a person of whom it may be said that he is estranged from himself, or that he finds himself a helpless or a passive bystander to the forces that move him.[38]

It is, in the end, rather difficult to spell out what Frankfurt thinks "estrangement from oneself" is. Partly this is because his views are quite complex, and partly it is because a succession of articles on the matter reveals changes in his line of thought. In what follows I will explain this development in his views, assessing how well his changing formulations capture the concept of freedom of the will/autonomy.

The concept of personhood

---

[38] *Ibid* 22. On page 170 he explicitly uses the term 'autonomy' when he says that:

> ...[I]t is these acts of ordering and of rejection...that create a self out of the raw materials of inner life. They define the intrapsychic constraints and boundaries with respect to which a person's *autonomy* may be threatened even by his own desires. [my italics]

In setting out his account Frankfurt begins by explaining something about the conditions under which we use the concept of *personhood*, so as to explain why we commonly understand freedom of the will to be closely connected to personhood. The sense of the term 'person' he is concerned with is that particular sense which, he says, is 'designed to capture those attributes which are the subject of our most humane concern with ourselves and the source of what we regard as most important and most problematical in our lives'.[39] And, as he believes that whether one can be described as a person or not depends on *what* one can desire - in other words he believes that one is a person only if certain states of affairs can be the object of one's desires - he offers a map of the different states of affairs which human beings can desire in order to clarify what it is that makes one a person.

The sense of the word 'person' which Frankfurt is concerned with is not the only sense of the word in common or philosophical usage - there is also the sense which serves merely to distinguish human beings from members of other species; and there is also the sense, employed in particular by the philosopher Peter Strawson, which serves to designate those beings to which '...*both* predicates ascribing states of consciousness *and* predicates ascribing corporeal characteristics...' can be applied.[40]

But, for Frankfurt, the term 'person' as a distinction between human beings and other species is philosophically uninteresting because it aims at distinguishing members

---

[39] *Ibid* 12.

[40] Strawson 101-2.

of our biological species from members of other biological species; and the sense used by Strawson, he says, is pernicious[41], in that it has contributed to the widespread neglect of what ought to be an important concern of philosophers - the concept of a person as capturing what creatures like us essentially are; the "deep" concept of personhood, if you like.

It is this "deep" concept of personhood that I will be commenting on below, and, as I will use it, it is *not* biologically defined. We tend to assume that only human beings can be persons, but the deep concept of personhood does not exclude the possibility that there may be other beings which are persons. And, if the concept is not biologically defined, there is no reason to assume that all human beings should be regarded as persons in this sense either.[42]

First- and second-order desires

The difference between persons and beings which are not persons, says Frankfurt, is that persons are capable of desiring certain states of affairs which non-persons are

---

[41] It does violence to our language to endorse the application of the term "person" to those numerous creatures which do have both psychological and material properties but which are manifestly not persons in any normal sense of the word.

Frankfurt 11.

[42] As we shall see, personhood, in the deep sense, is not usually attributed to children and certain kinds of mentally retarded adults. Frankfurt goes on to say on page 12 that '...these attributes would be of equal significance to us even if they were not in fact peculiar and common to the members of our own species'.

incapable of desiring. The desires which only persons are capable of having he calls 'second-order' desires - desires for other desires[43] - as contrasted with 'first-order' desires, which are desires to perform certain actions.[44] We not only desire to do certain things, as indeed do a great variety of creatures, but we are capable of desiring '...to be different, in [our] preferences and purposes...'[45] from what we are - a quality which is, according to Frankfurt, unique to us.

But setting out the distinction between first-order and second-order desires by describing first-order desires simply as *desires to perform certain actions* and second-order desires as *desires to have certain desires of the first order* may lead to confusion, says Frankfurt. This is because such descriptions don't specify the relative strengths of the desires that an agent may have at a given moment.[46] One may be able to say truthfully that one wants to perform a certain action, while, at the same time, desiring more strongly *not* to perform that selfsame action. The same could be said for second-order desires, although Frankfurt doesn't make this particular point: one might claim sincerely to want to have different desires from those which one does have, while wanting yet more strongly to retain the old set of desires.

---

[43] Or desires not to have certain other desires. In what follows I will take it as implicit that second-order desires can also be desires *not* to have certain other desires.

[44] Or not to perform them. In what follows I will also take it as implicit that first-order desires can be desires not to perform certain actions.

[45] Frankfurt 12.

[46] On page 13 Frankfurt specifies a whole number of statements which are consistent with the statement 'A wants to X'. Examples are 'A is unaware that he wants to X', 'A believes that he does not want to X', 'A wants to Y and believes that it is impossible for him both to Y and to X', amongst others.

It is necessary, therefore, when talking of desires, to specify whether or not one is referring to those desires which are *effective* - those desires which actually move one to action. This serves to clarify whether, when one states that A wants to X, one wishes to convey that this is the desire which is the strongest amongst the many first-order desires A may have. Such desires Frankfurt wishes, in setting up his account, to refer to as the agent's *will*[47].

A's will is, then, according to Frankfurt, one of A's first-order desires, namely A's *strongest* first-order desire. In this way Frankfurt distinguishes between his notion of the will and his notion of *intention*. As he sees intention, an agent may intend to perform a certain action, yet ultimately refrain from performing it because his or her intention is not as powerful as a conflicting desire which he or she also has.[48]

This classification of the kinds of desires creatures may have is complicated further by the fact that second-order desires may have two distinct objectives. In the first place one may want simply to have a particular first-order desire, without wanting that desire to be effective. Frankfurt's example of someone in this position is a physician who wishes to understand what it is like to desire a particular drug[49] - perhaps because he feels that he would be better able to understand addiction if he had some experience of the desire

---

[47] To identify an agent's will is either to identify the desire (or desires) by which he is motivated in some action he performs or to identify the desire (or desires) by which he will or would be motivated when or if he acts.

Frankfurt 14. Velleman uses the term "operative motive" for Frankfurt's "will". Cf (1992) 101 *Mind* 471 *et passim*. I shall adopt Frankfurt's usage of the term "will" hereafter, unless specified otherwise.

[48] Frankfurt 14.

[49] *Ibid* 14-5.

for the drug. But the physician does not actually want now to *take* the drug - in fact he takes steps to ensure that he won't be able to take it when the desire for it comes over him. In such cases the fact that A wants to want X does not imply that A actually wants X.

On the other hand, one may want a particular first-order desire to be the one that moves one to act - ie one may desire that a particular first-order desire be more powerful than one's other first-order desires.[50] Second-order desires of this kind Frankfurt terms 'second-order volitions'.[51]

It is important to note that Frankfurt's distinction between second-order volitions and other second-order desires does not mirror, at the higher level, his distinction between the will and other first-order desires. The will differs from other first-order desires in terms only of strength - it is the desire[52] which is powerful enough to move the agent to action. The analogous difference at the higher level would be the difference between those second-order desires which are actually fulfilled from those which are not - fulfilment being defined in this way: that a second-order desire of mine is fulfilled when I am actually moved to act on a desire which I desire to move me to act, and fails to be

---

[50] No doubt this situation is more common than that of the physician who wants to desire a drug without wanting the desire to be his will.

[51] Frankfurt 16.

[52] Or group of desires - the first-order desire which moves the agent to action may on its own be weaker than other first-order desires, but be able to move the agent to action because of added impetus from a second-order desire.

fulfilled when my will does not conform to the will I want.[53] This distinction is, however, *not* the same as Frankfurt's distinction between second-order volitions and other second-order desires. For him, second-order volitions differ from other second-order desires not in terms of being fulfilled, but rather in terms of whether the second-order desire is or is not pertinent to the identification of the agent's will. According to Frankfurt, an agent with a second-order volition desires that a certain first-order desire be paramount amongst his or her desires, whereas an agent whose second-order desire is not a second-order volition simply wants to have a particular first-order desire, without wanting to act on it.

So, to sum up, Frankfurt has outlined four different kinds of desire that a human being may have. Firstly: the class of first-order desires, which includes all desires to perform certain actions. Falling within this class are those desires which can be called the agent's will - those first-order desires which actually move the agent to action. Secondly: the class of second-order desires, which includes all desires to have or not to have particular first-order desires. Falling within this class are those desires which Frankfurt terms 'second-order volitions' - desires that particular first-order desires be those which move one to action. I leave aside for the moment the issue of desires of an order higher than the second.

---

[53] ...(T)he question of whether or not his second-order desire is fulfilled does not turn merely on whether the desire he wants is one of his desires. It turns on whether this desire is, as he wants it to be, his effective desire or will.

Frankfurt 16.

The context indicates that Frankfurt is referring here to second-order *volitions*.

Second-order volitions, personhood and freedom of the will

Frankfurt's motive for emphasizing that second-order volitions are distinct from simple second-order desires is that he considers the former and not the latter essential to personhood.[54] Creatures which lack second-order volitions he refers to as wantons. The difference between a person and a wanton can be summed up by saying that a person cares about his or her will, whereas a wanton doesn't. Wantons are not concerned with the question of whether or not they want to have the will they happen to have. They are happy to follow whichever path of action their strongest first-order desires lead them along, not caring which of their first-order desires is the strongest. Examples of wantons include, according to Frankfurt, '...all nonhuman animals that have desires and all very young children.'[55]

This appears to imply that acting with indifference to one's will is necessarily pernicious, or at least evidence of wantonness. But I would argue that there are certain clashes between lower-order desires towards the resolution of which one might legitimately be indifferent - particularly desires which one may act on without significant

---

[54] In fact, Frankfurt is a little ambiguous on this point. On page 12 he mentions that people '...are capable of wanting to be different, in their preferences and purposes, from what they are.', which implies merely that we have second-order desires, but not necessarily that we have second-order volitions. In fact he says explicitly, earlier on the same page, that 'It seems peculiarly characteristic of humans...that they are able to form what I shall call "second-order desires"...'

However, on page 16 he says this:

> It is logically possible, however unlikely, that there should be an agent with second-order desires but with no volitions of the second-order. Such a creature, in my view, would not be a person.

[55] *Ibid* 16.

consequences. For example, it would seem that being indifferent about whether my desire to spend an evening with friends or my desire to read a book eventually moves me to action hardly condemns me to wantonness.

One response to this suggestion Frankfurt might consider would be to say that wantonness is not necessarily pernicious. But I believe a better response would be to suggest that in the kind of case discussed above I cannot truly be described as a wanton because I have *taken up* an attitude of indifference to the resolution of the conflict between my lower-order desires, whereas a wanton simply has *no* attitude towards the outcome of the conflict. In this case we would need an explanation of what the difference between taking up an attitude of indifference and simply not caring is.[56]

In any event, Frankfurt goes on to say that adults may act more or less wantonly in connection with first-order desires they have, in that they may act with more or less indifference towards certain conflicts between their first-order desires. From this one may conclude that personhood and wantonness are to some extent a matter of degree.

Being a person, for Frankfurt, is a matter simply of caring about what will one has. One does not need in fact to *have* the will one wants to have in order to be a person. Frankfurt's example of a person who does *not* have the will they want is of an unwilling drug addict. This person has conflicting first-order desires - the desire for the drug and the desire to refrain from taking the drug (ie an aversion to the drug) - and wants (has a

---

[56] Frankfurt gestures towards a solution when he makes the following remarks on page 18:

> It would be misleading to say (of a wanton) that he is neutral as to the conflict between his desires, since this would suggest that he regards them as equally acceptable. Since he has no identity apart from his first-order desires, it is true neither that he prefers one to the other nor that he prefers not to take sides.

second-order desire) the desire to refrain from taking the drug to be the one which prevails. However, as it turns out more often than not, the desire to take the drug proves to be the stronger. But this is a cause of distress in the unwilling addict, indicating that he or she *cares* about what their will is.

It seems that Frankfurt would also have to regard *willing* addicts as persons, because *prima facie* willing addicts fulfil the condition for personhood in that they care about which desire moves them to act: willing addicts want the desire to take the drug to be their will. It is true that Frankfurt says that willing addicts lack freedom of the will because their desire to take the drug would be effective even if they *didn't* want the desire to constitute their will.[57] But this, in itself, does not seem to be sufficient to disqualify them from enjoying personhood. Certainly, the fact that they lack freedom of the will did not disqualify unwilling addicts from being considered persons.

Wanton addicts, on the other hand, would either take the drug or refrain from taking

---

[57] According to Frankfurt, the willing addict lacks freedom of the will in that the coincidence between such a person's will and his or her second-order volitions is *not their own doing*. This coincidence would be a matter of 'happy chance' - cf Frankfurt 20. He goes on to say, on page 24, that

> [a] persons will is free only if he is free to have the will he wants. This means that, with regard to any of his first-order desires, he is free either to make that desire his will or *to make some other first-order desire his will instead*. [my italics]

- a condition which the willing drug addict does not satisfy.

It is not clear whether Frankfurt means by 'some other first-order desire' some other first-order desire the agent already has, or whether he means simply any possible first-order desire. The former is far more plausible, but raises a puzzle about cases where the willing drug-addict experiences no desire to refrain from the drug whatsoever. There is then no alternative desire - alternative to the desire to take the drug - which the addict might make his will, if he so desires. Would such an addict enjoy freedom of the will?

Finally: The idea that the coincidence between someone's will and their second-order volitions could be not their own doing requires some elucidation in the light of these comments which Frankfurt makes on page 22:

> Examples such as the one concerning the unwilling addict may suggest that volitions of the second order...must be formed deliberately...But the conformity of a person's will to his higher-order volitions may be far more thoughtless and spontaneous than this...The enjoyment of freedom comes easily to some.

it according to which whim struck them in the moment, and, more importantly, *this would not strike them as a problem*.[58] 'When a *person* acts, the desire by which he is moved is either the will he wants or a will he wants to be without. When a *wanton* acts, it is neither.'[59]

Frankfurt sees second-order volitions as important also in that the capacity for forming them is linked to the capacity for enjoying *or* lacking freedom of the will, a capacity Frankfurt says has '...been considered a distinguishing mark of the human condition' and is 'essential to persons'[60].

Whereas enjoying freedom of action, he says, means (generally speaking) being able to do what one wants to do, enjoying freedom of the will means being able to have the will one wants to have, and, conversely, lacking freedom of the will means having a will one does not want to have. Whether one's actions are motivated by a desire by which one wants be motivated, or whether one's actions are motivated by a desire by which one does not want to be motivated, by definition, one *has* second-order volitions, because a second-order volition is a desire for a particular will.[61] Freedom of the will or, in my terms, autonomy, can only be possessed or fail to be possessed by beings that have second-order volitions.[62]

---

[58] *Ibid* 17.

[59] *Ibid* 19.

[60] *Ibid.*

[61] *Ibid* 20.

[62] *Ibid* 19.

Perhaps the best way of explaining this is to say that beings who don't care what their wills are - wantons, in other words - can't have freedom of the will for the simple reason that *there is no will that they want*. And, similarly, they can't lack freedom of the will either - because there is no will that they don't want. They can experience neither discrepancy nor harmony between their first-order desires and their second-order volitions because they do not have any second-order volitions.

The regress problem

Lacking the capacity to form second-order volitions is not the only cause, though, of wantonness, according to Frankfurt. This may also happen if agents experience conflict *between* their second-order volitions. It may be that, for example, agents are unable to decide which first-order desire they want to be their will because they have conflicting second-order volitions and they can't decide which of *these* desires they prefer. Such a situation also results in wanton behaviour - because if such agents act at all their will operates without their participation.[63]

It might seem that the obvious thing an agent in this situation could do to avoid

---

[63] *Ibid* 21:

> If there is unresolved conflict amongst someone's second-order desires, then he is in danger of having no second-order volition; for unless this conflict is resolved, he has no preference concerning which of his first-order desires is to be his will. This condition, if it is so severe that it prevents him from identifying himself in a sufficiently decisive way with *any* of his conflicting first-order desires, destroys him as a person. For it either tends to paralyse his will and to keep him from acting at all, or it tends to remove him from his will so that his will operates without his participation.

wantonness would be to form another, yet higher-order volition. In other words, the introduction of a third-order volition about the conflicting second-order volitions - wanting to have the one second-order volition rather than the other - might appear to be the solution.

It is not clear, however, that the introduction of a third-order volition to reinforce one of the conflicting second-order volitions is a genuine solution. Avoiding wantonness means, for Frankfurt, being clear on the question of what will one wants. But the introduction of a volition of a higher order brings us no nearer to answering that question. This is because, as Gary Watson points out,[64] there could be a conflict at that level as well, in which case we have no particular reason to regard any *third*-order desires as being pertinent to the identification of the agent's will.[65] And, as being in a state of conflict about one's will is one characteristic of wantonness, it appears that agents who have third-order desires reinforcing their second-order desires are as capable of being wanton as agents whose conflicting second-order desires are not reinforced by any third-order desire.

In fact, as should be clear from Watson's point, even if an agent doesn't have

---

[64] On page 218 of 'Free agency' :

> Since second-order volitions are themselves simply desires, to add them to the context of conflict is just to increase the number of contenders; it is not to give a special place to any of those in contention.

This passage refers explicitly only to the fact that we have no reason to regard *second*-order volitions as any more constitutive of the agent's will than any other desires, but the point is equally pertinent to desires of any order.

[65] The fact that the agent experiences no conflicting third-order desire is of no help. The mere fact of being unconflicted cannot establish a desire as being genuinely attributable to an agent, because, if this were so, unconflicted *first-order* desires which were not endorsed by desires of any higher order could then be considered genuinely attributable to the agent.

conflicting desires, we may still ask why that agent should not be considered to be wanton. The problem can be outlined as follows: If a creature who does not care which of its first-order desires moves it to action is to be considered wanton, why shouldn't wantonness be a feature of creatures who don't care whether their *second*-order volitions conform to yet higher-order volitions? And if second-order desires do need to be reinforced by higher-order volitions to avoid wantonness on the part of the agent, at which level can one be satisfied enough to look no higher? And how can any such termination point avoid arbitrariness?

Frankfurt argues that the infinite regress which appears to loom here can be avoided, for the reason that agents may identify themselves *decisively* with one or more of their first-order desires, and, in so doing, rescue their second-order volitions from arbitrariness. This is because, when they make such a commitment, the 'commitment "resounds" throughout the potentially endless array of higher orders'.[66] A commitment of this kind can be taken to mean that questions about higher-order desires simply don't arise for agents who have so committed themselves.

Clearly, however, Frankfurt came to feel unhappy with this proposed solution to the problem of the infinite regress, because in later papers he attempts not only to amplify the notion of "decisive identification" but also to sever it from the hierarchical model of first- and second-order desires which is at the heart of 'Freedom of the will and the concept of a person'.

---

[66] Frankfurt 21.

External desires

In one of these later papers, 'Identification and externality' Frankfurt begins by discussing the question whether the distinction between desires which are fully attributable to an agent and desires which are not (which he now calls 'external desires') is a genuine distinction - ie whether the idea of such a distinction really makes sense. In this context Frankfurt discusses the view of Terence Penelhum, who at one time believed it to be morally dubious to claim to be a victim of external desires.[67] Making such a claim, he said, '...denies that some desire...is part of one's ongoing history when it is'.[68] It is a strategy which is used in order to evade responsibility for one's actions. As Penelhum puts it, '...every desire must...belong to *someone*, and a desire with which a person does not identify himself clearly does not belong to anyone else.'[69]

Against Penelhum, Frankfurt offers an analogy between mental life and bodily activity. When speaking of the body, we routinely make a distinction between actions which are performed, and events which merely 'take place' in the body. Hitting someone, on the one hand, is a clear case of action. Experiencing a muscle spasm, on the other, we do not construe as an action, although it is an event which takes place in the body.

---

[67] 'The importance of self-identity' 670.

[68] *Ibid* 671.

[69] *Ibid* 670.

In other words, one may be active or passive with regard to the events which are part of one's ongoing physical history.

He suggests that analogous processes may be viewed in the life of the mind.[70] An example of an active psychological process would be '...turning one's mind in a certain direction, or deliberating systematically about a problem...'[71] But we can also be passive with regard to events which are part of our psychological histories. This happens when, for instance, we experience obsessional thoughts. Frankfurt argues that Penelhum's objection - that apparently external desires are just as much desires of agents as any other desires because these external desires can't be attributed to any *other* agent - fails to take into account the fact that we routinely distinguish between events in the history of agents according to whether the agent is active or passive in respect of that event. Frankfurt concedes that, of course, a desire I have must in a *literal* sense be my desire - but he claims that, within the class of those desires which are literally mine, there is a legitimate sense in which we say that some of those desires are not attributable to me.[72] We only attribute a mental event unequivocally to agents when we view them as active participants in it. When we view them as passive spectators we do reserve a sense in which the event cannot be attributed to them.

It is easier to think of examples of this if one looks at actions, as opposed to

---

[70] Frankfurt 59.

[71] *Ibid.*

[72] *Ibid* 61.

desires. Frankfurt's example is of someone in a crowded vehicle being pushed against a second person as a result of the vehicle's movements.[73] If the second person then asks who pushed them, while there is a sense in which it is true to say that the first person pushed them, it is also true that we would not regard a response of 'no-one pushed you' as moral evasion. It seems natural for us to say, in these circumstances, that there is no *action* which can be attributed to the first person, even though pushing the second person is clearly an event which took place in their physical history. Frankfurt wants to say that, analogously, we can make a similar claim about desires - that they can be events in the history of our minds without being attributable to us.

Thus, though Penelhum is correct in suggesting that it is morally dubious to deny that an event is part of one's ongoing history when it is, he is wrong in suggesting that this is what happens when someone claims to be the victim of an external desire. In making the claim one does not deny that the external desire is part of one's history - one denies rather that it can be attributed to one in anything other than a literal sense. More particularly, one denies that it can be attributed to one in the way that one's *actions* can be. This leaves open the possibility that it can be attributed to one in the way that involuntarily pushing another person in a moving vehicle can, a possibility that clearly acknowledges that the desire has a place in one's history.

No doubt such denials could very often simply be moral evasion. But, given that we are able to live with the possibility of moral evasion in cases where people claim that

---

[73] *Ibid* 60.

events in their physical history are not actions attributable to them, there appears to be no reason why we can't live with the same possibility in analogous cases in people's mental lives. Refusing to do so would be to throw the baby out with the bath water.

So it appears that, if one accepts the analogy between passivity with respect to events in one's physical history and passivity with respect to one's psychological history, there is, as Frankfurt puts it, '...a legitimate and interesting sense in which a person may experience a passion that is external to him, and that is strictly attributable neither to him nor to anyone else.'[74] But outlining exactly what it is that makes a particular desire external proves to be difficult. At first glance, there appear to be a number of promising possibilities for conditions of externality.

Conditions of externality

(I) Artificial inducement

Firstly, it might seem likely that desires which are artificially induced would be external. Desires implanted as a result of hypnosis, or the use of drugs, might appear to be, on the face of it, paradigm cases of external desires. Frankfurt, however, is reluctant to accept

---

[74] *Ibid* 61.

that artificial inducement necessarily implies externality. This is because, he says, although such desires present themselves to the person who experiences them as '...discontinuous with his understanding of his situation and with his conception of himself'[75] people are able to rationalize these discontinuities, convincing themselves that the apparently external desires they are experiencing actually do have a legitimate place in the set of their desires. When this happens, says Frankfurt, they can no longer claim to be passive with respect to these desires.[76]

(ii) Irresistibility

Irresistibility might also appear to be a sufficient condition of externality. One might argue, for example, that the desire willing drug-addicts feel for the drug to which they are addicted is an external desire, as they would be incapable of resisting it should they so wish.

But, for Frankfurt, a willing drug-addict is actually in the same position as someone who rationalizes an artificially induced desire. Although it's likely that people in such a position would be unable to resist their desires, they might nevertheless endorse them - ie regard them as internal. Likewise, according to Frankfurt, the desire willing drug-addicts have for the drug is their *own*, because their first-order desires for the drug are

---

[75] *Ibid* 62.

[76] *Ibid.*

endorsed by second-order desires[77] - which implies that the desire for the drug is thereby made internal. As with artificially induced desires, then, one may convince oneself that an irresistible desire does have a legitimate place in the set of one's desires, and then no longer be entitled to claim that one is passive with respect to that desire.

Frankfurt goes on to say that the belief that irresistibility is essential to externality arises because people generally only bother to disclaim desires they experience when they are unable to resist them. But, in fact, it is quite possible that one should have a desire with which one doesn't identify without succumbing to it.[78]

(iii) Disapproval

The third possible candidate for a condition of externality is disapproval. It seems quite natural to say of those desires of ours of which we disapprove, when we do experience them, that they are external or alien.[79] This is because we tend to experience them as a violation of some kind, or as being inconsistent with those desires we believe are genuinely attributable to us.

Frankfurt points out, however, that we should be careful not to take disapproval as a sufficient condition of externality, as it is possible that one may eventually become

---

[77] *Ibid* 20. He says nevertheless that they do not enjoy freedom of the will.

[78] *Ibid* 64.

[79] Gary Watson makes a similar point on page 72 of 'Free agency' when he suggests that unfree action is action motivated by desires which we do not *value*.

resigned to certain character traits or desires which one thinks of as defects and therefore disapproves of. In cases such as this, although one retains one's disapproval of the desire in question, in that one regards it as a defect in one's character, one accepts it as internal, as an integral part of one's personality.[80]

He also goes on to say that although attitudes of approval and disapproval appear to have something to do with externality and internality - because it is difficult to think of a case in which someone to whom a desire is external nevertheless approves of it[81] - it is impossible to explicate the concepts of internality and externality in terms of attitudes. The reason is that any such attempt ignores the fact that, as Frankfurt puts it, 'attitudes towards passions are as susceptible to externality as are passions themselves'.[82] The point is that the fact that one disapproves of a desire cannot show that desire to be external because the question still arises as to whether one's disapproval is a genuine expression of one's self as opposed to an external imposition. Likewise for approval. If one claims that a desire one experiences is internal on the grounds that it is a desire one approves of, one can be required to show that one's approval is internal.[83] The infinite

---

[80] Frankfurt 63-4. John Christman is unhappy with the implication that desires which one disapproves of but is resigned to can then be considered internal desires. According to him:

> Even the desires that are the result of obviously heteronomous processes can be viewed as being a (regrettable) part of oneself, maybe something one cannot change and for a time something one is simply 'stuck with.' In this way I can just as readily 'identify' with those *non*-autonomous aspects of myself as the more 'authentic' parts.

Cf Christman (1991) 21 *Canadian journal of philosophy* 5.

[81] Frankfurt 65.

[82] *Ibid* 65.

[83] This is not the only reason for rejecting the notion that autonomy requires acting only on those desires which one approves of. John Christman points out on page 5 of 'Autonomy and personal history' that '...to be autonomous in this way, I would have to be in some sense

regress we met in 'Freedom of the will and the concept of a person' looms once more.

(iv) Decisive rejection

At this point Frankfurt confesses his inability to offer a clear characterization of what it is that makes a desire internal or external.[84] Towards the end of 'Identification and externality' he returns, however, to the notion of decisive identification as offering the rudiments of a solution when he points out that people experience two different kinds of conflict of desire.[85]

Typical of the first kind is the fact that if circumstances make it impossible to satisfy the strongest of the competing desires, one will then attempt to satisfy the desire which ranks next in terms of strength. The example Frankfurt gives is of a conflict between one's desire to see a film and one's desire to go to a restaurant for the evening.[86] One may, for example, decide that the desire to go to a film is the stronger one, and therefore make the effort to satisfy it. If, however, it becomes impossible for some reason to see a film, the natural thing to do would be to go to a restaurant instead. In this example, the path to solving the conflict involves *ordering* the conflicting desires.

---

perfect...'

[84] Frankfurt 65.

[85] *Ibid* 66 *et seq.*

[86] *Ibid.*

The second kind of conflict is solved not by ordering the conflicting desires, but by *rejecting* the one desire.[87] This means that if one is prevented from satisfying the one desire one is not led to try to satisfy the other. Frankfurt offers as an example of this kind of conflict the case of a man who wants to compliment a friend on a recent achievement but who also feels inclined to injure the friend out of spite.[88] If he decides that he is going to compliment the friend, but is in the end only presented with an opportunity to injure him, he would not naturally take such an opportunity. This is because, as Frankfurt puts it, the two desires belong to different orderings.[89]

In this second kind of conflict one cannot say, as one can in the case of the first kind of conflict, that one wants to compliment the friend *more* than one wants to injure him. There is a sense in which one doesn't want to injure him at all, even though one experiences the desire to injure him. And if the desire to compliment the friend wins out, one might say not that it was the stronger of the two desires, but rather that the '...*person*, who wants to pay his acquaintance a compliment is stronger than the desire to injure him that he finds within himself.'[90] This suggests that it is the *decision* to reject the one desire which makes it external.

---

[87] *Ibid* 67.

[88] *Ibid* 66-7.

[89] Taylor makes a similar point on page 283 in 'Responsibility for self' when he says that :

> ...[I]n non-qualitative reflection one desired alternative is set aside...only on grounds of its contingent incompatibility with a more desired alternative. But with qualitative reflection this is not necessarily the case. Some desired consummation may be eschewed not because it is incompatible with another, or if because of incompatibility, this will not be contingent. Thus I refrain from committing some cowardly act, although very tempted to do so, but this is not because this act at this moment would make any other desired act impossible, but rather because it is base.

[90] Frankfurt 68.

It may be, he says, '...that a decision...lies behind every instance of the establishment of the internality or externality of passions'[91]. Or, he continues, '...perhaps it is by referring to something more general, of which decisions are only special cases, that we must seek to understand the phenomena in question'[92].

Values and autonomy

In response to the difficulties of Frankfurt's version of the subjective view, Gary Watson offers an alternative. Like Frankfurt, he wishes to explain what it is that distinguishes autonomous agents, whom he refers to as 'free agents'[93], from those who are not, and his view shares with Frankfurt the characteristic of regarding agents' own assessment of their desires as crucial in ascertaining whether those desires are autonomous. [Maybe this is the crucial problem - if there was to be some sort of objective assessment of which values really belonged to an agent then perhaps conflicting desires *could* be genuinely attributed to the agent, but as long as the agent him/herself is supposed to be the arbiter then we will want to know *from which standpoint* the agent is judging] However, he differs from Frankfurt in that he sees autonomous desires as those which conform to agents'

---

[91] *Ibid.*

[92] *Ibid.*

[93] Cf the title of his paper - 'Free agency'.

evaluations, whereas, as we have seen, Frankfurt's view in 'Freedom of the will and the concept of a person' is that autonomous desires are those which conform to the second-order volitions of agents.

Watson's suggestion is that we ought to look at the peculiar *quality*[94] of the desires in question in order to distinguish the autonomous from the heteronomous. As I have mentioned above, Watson argues that Frankfurt's notion of orders of desires cannot tell us '...why or how a particular want can have, among all of a person's desires, the special property of being peculiarly his "own".'[95] One may be alienated from one's second or higher order desires and it is therefore always a legitimate question as to whether such desires are in fact genuinely attributable to one. The fact that they belong to a higher order is not enough to *make* them autonomous desires. Watson accepts that Frankfurt may be correct in thinking that the desires which can genuinely be attributed to agents may be constituted by acts of identification or by decisive commitments, but points out, correctly, in my opinion, that if these are the crucial notions then there is no need to talk about hierarchies of desire.[96]

Frankfurtian hierarchical maps of desires have not only drawn criticism for the logical problems associated with them, but also for their alleged empirical distortions.

---

[94] I owe this characterization of the difference between Frankfurt's early approach and that of Gary Watson to Susan Wolf. Cf *Freedom within reason* 30.

[95] Watson (1975) 72 *Journal of philosophy* 218.

[96] *Ibid* 219. He continues to say that if acts of identification or decisive commitment are what constitute our real desires '...no ascent is necessary, and notion of higher-order volitions becomes superfluous or at least secondary.'

Watson argues that our volitional lives do not necessarily involve endorsing or rejecting already existing first-order desires. He points out that our reflective assessments may induce in us desires that we had not hitherto experienced.[97] And he goes on to say that we are not so much concerned with which desires motivate us as we are with the question of which course of action is best.[98] This point is also made by Irving Thalberg, who says that our '...second-order volitional antics are more concerned with our behaviour, and its effect, than with the first-order desires that engendered it.'[99]

According to Watson, when we say that a desire we have is not fully ours, we mean by this that it is a desire for something which we do not *value*.[100] Underlying this claim is a distinction, originating in Plato's conception of different parts of the soul,[101] between our motivational systems and our valuational systems - a view which sees desires which are the agent's own and those which are not as originating from different sources, rather than as occurring on different levels in a hierarchical structure. Our motivational system can be likened to Frankfurtian first-order desires - it is '...that set of considerations which move [us] to action.'[102] Now although these desires are literally ours, there is also a sense in which they can sometimes be considered not really attributable to us. This is possible

---

[97] *Ibid.*

[98] *Ibid.*

[99] He also asks on page 215 whether '...we need [first- and second-order conation] to interpret even...Frankfurt's hand-picked cases.'

[100] Watson (1975) 72 *Journal of philosophy* 216 *passim.*

[101] *Ibid* 207 *et passim.*

[102] *Ibid* 215.

when they motivate us in spite of our valuational systems, which Watson describes as '...that set of considerations which, when combined with [an agent's] factual beliefs...yields judgements of the form: the thing for me to do in these circumstances, all things considered, is *a*.'[103] Autonomous action, then, for Watson, would issue from an agent who acts according to his or her valuational system.

Susan Wolf has suggested that Watson's conception of values needs to be expanded somewhat.[104] This is because, according to her, it makes one's values too heavily reliant on one's reason. As outlined above, Watson takes our values to be judgements. Wolf, however, believes that it would be more accurate and inclusive to regard anything that we care about as our values, given that many people may have emotional, or otherwise nonrational, commitments to particular values which could not be expressed in the form of judgements, or even necessarily be justified.

Such a broadening of the theory does not affect Watson's central claim, which is that the advantage of regarding one's values as the desires which are one's own is that one cannot dissociate oneself from one's valuational system.[105] And, clearly, if this were true, his account would not face the infinite regress problem that Frankfurt's hierarchical account does - for there would be no need to ascend to ever higher levels of identification. It is important to note that when Watson says that one cannot feel

---

[103] *Ibid.*

[104] Wolf *Freedom within reason* 31.

[105] Watson (1975) 72 *Journal of philosophy* 216.

alienated from one's values he is not claiming that one cannot feel alienated from particular values that one might hold. One clearly can be, in that the question might arise as to whether one really identifies with that particular value. His claim is that '...one cannot coherently dissociate oneself from [one's valuational system] *in its entirety*.'[106]

Unfortunately, however, it is not at all clear that this claim is true, and Watson's apparent solution to the regress problem is not as successful as it might appear to be. This is because, in fact, just as one may feel alienated from a particular value one holds, one may also feel alienated from a system of values. Just as one may not be sure that a particular value one undoubtedly holds is one's *real* value (perhaps because one is not sure whether it is worth valuing), one may also be uncertain about an entire set of values. It is easy enough to imagine such a condition: Anyone who moves between two or more cultures or subcultures may become confused about their identity, and may wonder which *system* of values is genuinely theirs. Watson appears to anticipate such problems when he says that:

> ...it does not follow from the fact that one must assume a standpoint that one must have only one, nor that one's standpoint must be completely determinate. There may be ultimate conflicts, irresolvable tensions, and things about which one simply does not know what to do or say.. Some of these possibilities point to problems about the unity of the person.[107]

---

[106] *Ibid.*

[107] *Ibid.*

But this seems to me to gloss over the issue. If one's freedom arises from the ability to act in accordance with one's values, serious questions for that freedom arise if one in fact is not sure what one's values are. How is one to know which system is in fact the location of one's *real* values? It is hard to see how the substitution of values for desires, and the substitution of independent sources of motivation for hierarchies of desire solves this problem. If there are serious enough conflicts amongst the agent's values, as Watson concedes is possible, then we face a parallel problem to the regress Frankfurt faced. It will not do as an answer to the question of which values are the agent's real values, as J David Velleman has pointed out,[108] to respond that those values which are not merely lodged in the agent but are actually fully integrated into the agent's valuational system are the agent's own desires, because we are then still left with the question of how values are integrated into this system, and how we know which ones are and which ones aren't. Must the agent identify with these values for them to be fully integrated? If so, how does this take place? By endorsement from a 'higher-order value'? These are precisely the problems Frankfurt's hierarchical account faced.

Christman points out[109] that there are two different senses of the term "identification". One could understand it to mean simply *acknowledgement* that one has certain desires or values. This sense, however, would not be of any help in assessing which of the agent's values his or her actions must conform with to be considered

---

[108] Velleman (1992) 101 *Mind* 461-81footnote 26.

[109] (1991) 21 *Canadian journal of philosophy* 5-6.

autonomous, because it is quite possible for one just to acknowledge regretfully that a desire or value is one's own despite not being autonomous with regard to it. The second sense Christman points out is one of "approval". In this sense one would identify with a desire if one approved of its existence within one's value system. But, as he points out, if identification with one's values in this sense were required for autonomy, one would have to be perfect in one's own eyes to be autonomous - a clearly excessive demand.[110]

There is a further difficulty which Watson provides no more help in overcoming than Frankfurt. And that is, why should we regard those desires which conform to the valuational system in its entirety, in the case of Watson, or those which the agent endorses by means of a second-order volition, as in Frankfurt, as the autonomous desires of the agent? Watson himself argues, in connection with Frankfurt, that second-order volitions have no special claim to being the agent's real desires.[111] Why then should values be any more legitimate as contenders for identification as the real self of the agent than any other mental items? As Thalberg points out, this flies in the face of the commonly accepted Freudian vision of the self as made up of '...conflict-prone systems of libidinal, destructive, morbid, self-preserving, sociable, conscientious, guilt-ridden...forces...'[112] One does not have to be a committed Freudian to see that Watson has much that he needs to argue for on this particular point.

---

[110] *Ibid.*

[111] Watson (1975) 72 *Journal of philosophy* 218.

[112] Thalberg 225.

****

Chapter 3: Subjective views of autonomy: The later Frankfurt and Velleman

Decisive identification

Given the inadequacies of Watson's attempt to deal with the difficulties faced by the early Frankfurt, I wish now to turn to Frankfurt's later refinements of his position, which I believe offer a solution to the regress problem at least. In his paper 'Identification and wholeheartedness,' he fully recognizes the difficulty with the hierarchical model, which is, to recapitulate, that it is not clear why, although first-order desires need the endorsement of a higher-level volition before they can be said to be fully attributable to the agent, this stricture apparently does not apply to second-order volitions. After all, there is no reason to suppose that second-order volitions are any more attributable to the person than any other desires they have, whether they be of lower or higher orders. It may be just as plausible to suggest that one's *first*-order desires are properly attributable to one, and that any higher levels are effete refinements or, as Thalberg suggests, cowardly second thoughts.[113] Alternatively, one might suggest that one must retreat to levels higher than the second-order before one can properly identify with any particular desire.

---

[113] 'Hierarchical analyses of unfree action' 220.

To forbid the question of whether agents' second-order desires are fully attributable to them appears arbitrary. But, on the other hand, if we allow access to higher levels of desires, saying that lower level desires can only be attributed to the agent if they are endorsed by desires at a higher level, we face the regress problem. To avoid this, a point would have to be arrived at where the sequence was terminated. But, as we have seen, it is difficult to specify under what conditions this could be done without the termination seeming as arbitrary as simply specifying that desires of a certain order are as high as we are prepared to go.

The solution which Frankfurt gestured at in 'Freedom of the will and the concept of a person' and 'Identification and externality' is that when one *identifies decisively* with a particular desire, there is no need to ask whether this identification is endorsed by higher-order desires. In 'Identification and wholeheartedness', he admits that the notion of decisive identification is highly obscure and aims at refining and expanding his account of what happens when one identifies with a desire.

As a preliminary to an attempt to solve this problem, it is important to make explicit, although Frankfurt himself does not do so, that it is not necessary to insist that the identification must be with a desire of a higher order than the first. What is actually required is a demonstration of how a desire of *any* level can be taken as genuinely attributable to a person without having to be endorsed by a higher-order desire.

It may turn out that the majority of desires which can be shown to be attributable to a given person *are* desires of the first order. But it is equally possible that some of

them may be second- or third-order desires. Desires of a higher level than the third would indicate a rare degree of mental sophistication. But forming desires of extraordinarily high orders might be equally indicative of a neurosis which causes agents to be unable to identify decisively with any of their lower-order desires.

Now to Frankfurt's solution to the problem of how to call a halt to the threatened regress. According to Frankfurt, decisive commitment is not an arbitrary way of cutting off the ascent to higher orders of desire. When considering whether a desire is genuinely one's own or not, there usually (if not always) comes a point where it is possible to decide *rationally* that recourse to a higher level desire is not necessary. To make this point he offers an analogy which compares the situation of someone trying to decide whether to identify themselves with one of their desires with the situation of someone solving an arithmetical problem.[114]

When someone solving an arithmetical problem arrives at an answer to their calculations, they may then embark on a process of checking this answer. In doing so they may perform the same calculation again, or an equivalent one, and will usually reach a point where they are satisfied that they have the correct answer. Once they are satisfied in this way, they then terminate the checking process.

Although, theoretically, one might continue checking the results of one's calculation indefinitely, as there is nothing about any particular number in the sequence of checks

---

[114] Frankfurt 167-9.

which guarantees that the correct answer has been arrived at, there is a point beyond which it is unnecessary to continue checking. Such a point would be reached when the person doing the calculations comes to believe that any further calculations will yield the same result as the previous ones. Or, though they may feel somewhat less confident about the results of future calculations, they may believe nevertheless that further checks would not be worth the effort involved. Daniel Dennett makes a similar point when he says that 'it would clearly...be irrational to embark on a limitless round of self-evaluation...Moreover, there is almost certainly no "book" answer to the question of how much moderation is the right amount of moderation.'[115]

In forming judgements of this kind, the person doing the calculations makes a decisive commitment to a particular answer. They decide to 'stick with' that answer. In the first kind of case mentioned, the commitment is to the correctness of the mathematical solution arrived at, and in the second kind of case, the commitment is to the judgement that further calculations would not be worth the effort.

These decisive commitments 'resound' through further potential calculations, says Frankfurt, in that the person making the commitment either confidently expects future calculations to yield the same answer, or confidently expects that further thought would confirm that adopting the answer is the most reasonable alternative in the circumstances. There is therefore no reason to continue checking - as Dennett says, to do so would be irrational.

---

[115] *Elbow room* 87.

Frankfurt argues that checking the results of an arithmetical calculation is analogous to forming desires of higher orders. This is, he says, because people do these two things for the same reason - namely, to eliminate conflict.[116] People who check their calculations do so perhaps because their first and second attempts at the problem yielded conflicting answers, or because they simply think that they may have arrived at an incorrect answer.

Similarly, agents who form desires of higher orders do so because of conflict between desires at a lower level, or because they feel unsure about whether they really do endorse lower-order desires they have. In both cases, it is reasonable to stop when they believe that they have eliminated the conflict that was worrying them. People checking their calculations stop when they realize that further inquiry would not lead them to change their minds, and agents forming higher-order desires stop when they no longer see any conflict to be resolved.[117]

In neither case, says Frankfurt, would such a commitment be an arbitrary one. The person identifies with a desire or endorses a particular answer to the mathematical problem when refusing to do so would be to demonstrate a greater degree of arbitrariness than doing so would.

It is important to note, as Frankfurt does[118], that the point here is not that one

---

[116] Frankfurt 169.

[117] *Ibid.*

[118] *Ibid.*

cannot be mistaken about where to stop the enquiry. Obviously, a judgement about whether the right point for terminating the sequence had been reached or not would be a fallible judgement. What is being claimed, however, is that if such judgements have been reasonably arrived at, it would be rational to follow them. This is in contrast to the suggestion that it could never be rational to terminate the sequence at a given point. When someone identifies with a particular desire, in Frankfurt's model, what they do is to *constitute their self*. Identifying with a desire in such a way is to settle the question of which of two (or any number of) desires is the desire genuinely attributable to a person.

By identifying with one of the rival desires in a conflict, an agent changes the conflict from one which divides the self to one between those desires which are genuinely attributable to them and alien or external desires. This does not necessarily eliminate the conflict. As a matter of fact, the rejected desire may remain more powerful than its rival with which the person has decided to identify, but the rejected desire cannot then be genuinely attributed to the agent. Although in such cases the person may still not be able to *have* the will they want, at least they *know* which will they want, and can therefore be said to have an integrated self.[119]

If we now return to the question of the nature of autonomy or freedom of the will, it seems that the answer which the Frankfurt of 'Identification and wholeheartedness' would give is that we act autonomously when we act on a desire with which we have decisively identified.

---

[119] *Ibid* 173-4.

Alienated decisions

I argued in the previous section that Frankfurt solves the regress problem in 'Identification and Wholeheartedness', and that this account does not infinitely proliferate decisions. Unfortunately, however, it fails for another reason, namely that decisions, whether capable of being brought to a halt or not, are not the *kind* of mental item which *could* guarantee autonomous behaviour as their upshot. This is one of the points made by J David Velleman in his article 'What happens when someone acts?'.[120] He argues there that although Frankfurt is correct in claiming that autonomous action issues from desires with which the agent has identified, this claim does not advance our understanding of autonomous action in that it fails to offer an adequate explanation of what identification consists in. Identification is as mysterious a notion as the self, and is itself in need of explication.

Frankfurt's belief, as outlined above, is that a *decision* is what constitutes an agent's identification with a particular desire. The problem is, according to Velleman, that one can be alienated from a decision - a decision can simply take place in one without one playing an active role in it.[121] So identification cannot be *constituted* by a decision;

---

[120] (1992) 101 *Mind* 461-81.

[121] Dennett makes the same point on page 80 :

> ...[N]ote how many of the important turning points in our lives were unaccompanied, so far as retrospective memory of conscious experience goes, by *conscious* decisions. "I have decided to take the job," one says. And very clearly one takes oneself to be reporting on something one has done recently, but reminiscence shows only

and the fact that one has decided to identify with a desire is thus no guarantee that autonomous action will ensue

To illustrate this point, Velleman offers an example of someone meeting an old friend for the purpose of resolving a dispute between the two of them. The meeting has been agreed upon some time back. During the course of the meeting the first person's replies become sharper, in response to comments which his friend makes, resulting in the two of them parting in anger. On reflecting on the outcome of the meeting, the first person realizes that his rise in temper was brought on by grievances which he had been mulling over during the period leading up to the meeting. These grievances had, as Velleman puts it, 'crystallized in (his) mind...into a resolution to sever (their) friendship over the matter at hand...'[122].

It would be accurate to say of this incident, he argues, that his behaviour was caused by a decision of his. But he goes on to say that it cannot truthfully be said to have been a decision he *made* or *carried out*. It would be more accurate to say that the decision was 'induced in' him and not 'formed by' him, and that, although the decision was genuinely executed in his behaviour, it was executed without his help.[123] A similar point is made by John Christman.[124] He argues that the notion of decisive identification outlined by Frankfurt is fatally ambiguous and can therefore not serve as a guarantor of

---

that yesterday one was undecided, and today one is no longer undecided...

[122] (1992) 101 *Mind* 464.

[123] *Ibid* 464-5.

[124](1991) 21 *Canadian journal of philosophy* 8-9.

the autonomy of the agent who identifies decisively with a particular desire. Frankfurt's view is that one identifies decisively with a desire if one believes that no further accurate inquiry would require the agent to change his or her mind. But a decision of this kind must nevertheless be arrived at by an agent acting according to the information and reasoning capacities that he or she possesses at the time. If this is so, then, according to Christman, 'the possiblity exists that a thoroughly manipulated individual could be declared autonomous'.[125] He imagines a case where one is hypnotised into having a particular desire, where the hypnotist includes a direction to ignore any information concerning the hypnosis itself. In this case, the agent would, in all likelihood, "decisively" endorse the desire which had been implanted in them, and would then, if Frankfurt is correct, be autonomous. Yet the decision involved in such a case would clearly be external.

Because of their susceptibility to internality or externality, then, decisions cannot constitute identification; and so the fact that an agent decided to act in a certain way does not mean that the resulting actions can genuinely be attributed to him or her. The decision may have taken place without the agent's participation.

The desire to act according to reasons

[125] *Ibid* 9.

The notion of identification may still prove useful to a theory of autonomy if an accurate characterization of what it consists in can be outlined. Velleman believes that the desire *to act in accordance with reasons* serves just such a purpose.[126]

It is not possible, according to him, for agents either to identify with or be alienated from this desire, because the desire to act in accordance with reasons actually *constitutes* agency. He arrives at this conclusion by arguing that this desire - the desire to act in accordance with reasons - is *necessarily* the driving force behind practical thought itself.[127] In his words:

> We say that the agent calculates the relative strengths of the reasons before him; but in fact, these calculations are driven by his desire to act in accordance with reasons. We say that the agent throws his weight behind the motives that provide the strongest reasons; but what is thrown behind those motives, in fact, is the additional motivating force of the desire to act in accordance with reasons.[128]

Reasons cannot literally drive us to act. We undertake what seems to us the most reasonable course of action only if we have a desire to act in accordance with reasons. And if one did not have such a desire one would be, in Frankfurt's terms, wanton, because one would not care what one's motivations were. To suppress the desire to act

---

[126] ' (1992) 101 *Mind* 478.

[127] *Ibid* 478-80.

[128] *Ibid* 479.

in accordance with reasons, then, is to cease assessing one's motivations, and to do this would be to lose any identity apart from one's other desires.

For Velleman, autonomous action is action which includes the desire to act according to reasons amongst its determinants. Heteronomous action would be action undertaken in cases where the actor[129] was not, or was unable to be, motivated by the desire to act according to reasons.

Conclusion

I have canvassed a number of suggestions as to which desires can be genuinely attributed to agents. All of these views are subjective views of autonomy, in that they share in common the feature that they regard agents' own assessments of their behaviour as crucial in deciding whether they are autonomous or not. These views face a number of problems.

Firstly, any theory of autonomy which relies on agents' assessment of their own behaviour - in other words any theory which holds that agents themselves determine whether or not they enjoy autonomy - has to avoid a regress. Both Frankfurt and Watson go to great lengths to show that, in identifying a certain mental item as the spring of

---

[129] I use the term 'actor' here because Velleman's usage of the term 'agent' to mean a being which acts in accordance with reasons would preclude it from being used here.

autonomous action, they are not thereby committed to an infinite number of such items. I have argued that both the early Frankfurt's use of the notion of second-order volitions and Watson's use of the notion of the valuational system of an agent cannot circumvent this problem, as it seems that in order to be genuinely attributed to the agent, both second-order volitions and values need the further endorsement of yet further volitions and values respectively.

In Harry Frankfurt's later article 'Identification and wholeheartedness' he sketches an account of how decisive identification with one of one's conflicting desires can solve the regress problem. He does so by drawing an analogy between solving a mathematical puzzle and resolving conflict between one's desires. At some point, he argues, continuing to assess which desire one wishes to identify with is more irrational than calling a halt to the regress by identifying decisively with one of the desires. However, J David Velleman highlights the fact that identification with a desire can't consist in a decision to do so, as it is possible for one to be alienated from one's decisions. Autonomous action does not issue from decisive identification with one of one's desires, but rather, according to him, from the desire to act in accordance with reasons, as this desire is a mental item from which it is not possible to be alienated without losing one's status as an agent.

But it is not clear that Velleman's account is better than either Frankfurt's or Watson's at coping with the crucial problem for the subjective view. This is that the problem that, as

Bernard Berofsky puts it, 'the real self may be ill and, therefore, not responsible'[130]. We can think of cases where the desires of an agent may satisfy the requirements for autonomy as set out by Frankfurt and Watson, and his or her actions may be determined by the desire to act in accordance with reasons, as required by Velleman, and yet we would not regard the agent in question as autonomous or responsible for their actions. John Christman imagines the following case:

> ...a person who lives a completely subservient life and who also identifies with the first order desires that comprise such a life. Socialization and fierce conditioning throughout the person's life lead her to adopt, let us say, the life of complete subservience as her true calling. Thus, on the hierarchical analysis, she passes the test of autonomy since her higher-order desires are consistent with her lower-order desires. She approves of the lower-order desires, and identifies with them.[131]

Susan Wolf makes a similar point[132] when she refers to cases of people such as the so-called 'Son of Sam' murderer who, apparently, acted on the basis of reflective judgements and evaluations, but even so was clearly seriously mentally ill. It is quite possible that the Son of Sam satisfied both Watson's criterion of being motivated in conformity with his evaluations and Frankfurt's criterion of being motivated in conformity with his second-order volitions, and yet we would balk at the suggestion that he was an

---

[130] (1992) 89 *Journal of philosophy* 203.

[131] (1991) 21 *Canadian journal of philosophy* 6-7.

[132] *Freedom within reason* 37.

autonomous agent.

Furthermore, it is quite possible to think of examples where peoples' actions are informed by the desire to act in accordance with reasons, as required by Velleman, but whose actions nevertheless indicate that they are not fully in control of themselves. Denise Meyerson mentions the case of someone who works compulsively, believing that they have good reasons for doing little other than work.[133] She points out that our typical response to such a person is that they are 'driven' - an epithet which carries with it the implication that they are not fully in control of themselves. The mere presence of the desire to act according to reasons is not in itself a guarantor of autonomous action.

It therefore seems that subjective views of autonomy cover too much ground in that people such as the subservient woman, the Son of Sam and the workaholic could, on these criteria, be classified as autonomous. It may be that motives, values, reasons and actions must meet some objective standard, regardless of how they are viewed by the agent in question, before they can be considered autonomous.

I wish, in the following chapter, to assess whether what I will call the *objective* view of autonomy - the view that the beliefs and desires of an autonomous agent must satisfy some normative standard irrespective of the agent's appraisal of them - provides an account which succeeds where the subjective view fails, and further, whether it can provide us with a complete theory of autonomy.

---

[133] (1994) 54 *Analysis* 173.

****

Chapter 4 : Objective views of autonomy

The argument of the thesis up to this point has progressed as follows: If we wish to explain why we regard certain desires as being genuinely attributable to agents and others not, we need to be able to find the significant difference between these two kinds of desire. In the first chapter I dismissed the suggestion that desires which cannot be attributed to an agent arise as a result of influences external to that agent, while autonomous desires - those which are genuinely attributable to an agent - arise free from such influences. In the second and third chapters I discussed the view that an agent's autonomous desires or actions are those which the agent regards as internal in some specified way, regardless of whether or not they arose originally from external influences. This subjective view does coincide with our intuitive responses to certain classes of actions usually taken to be heteronomous[134], but there are also cases in which the subjective view is intuitively unsatisfying.

I think here of the cases mentioned by Susan Wolf[135] where we are disinclined to regard certain people as autonomous even though they undoubtedly identify with their

---

[134] It is easy, for example, to see how our intuition that kleptomaniacs are not responsible for their behaviour could be explained by the view that the desire to steal experienced by the kleptomaniac is an external desire. Likewise, hypnotized people could be understood to act on motives which are not their *real* motives.

[135] *Freedom within reason* 37.

motives.[136] Examples of such people which I mentioned in chapter two were the workaholic, John Christman's example of the woman who is socialized into a subservient life[137] and most cases of Son of Sam type murderers whose childhoods inhibited self-reflection on their desires. The fact that such agents identify with the motives from which their actions spring would nevertheless be unlikely to prevent us from regarding them as heteronomous agents. We would be likely to believe that their selves, from which their actions issue, were defective in some way. Neither Watson's understanding of autonomous action as action proceeding on the basis of one's values rather than one's brute desires, Frankfurt's understanding that it issues from desires with which the agent identifies or Velleman's view that autonomous action must be informed by the desire to act in accordance with reasons can accommodate the intuition that some actions are not autonomous because the *selves* from which the actions issue are in some sense heteronomous.

On the subjective view, being able to act in accordance with one's values, or being motivated by desires one identifies with or by the desire to act in accordance with reasons simply *is* autonomy. No further questions can be asked with regard to an agent's autonomy once one knows that the agent in question has identified with the desires or actions in the way required by the view one adheres to. Subjective views, then, assume that the notion of acting in accordance with desires one identifies with is sufficient in itself

---

[136]See, for example, Wright Neely's comment that '...freedom is not just a matter of doing as one desires, but requires, in addition, that we should have something to say about what we desire.' Neely 37.

[137](1991) 21 *Canadian journal of philosophy* 6-7.

to capture what it is to regard someone as autonomous, or to regard them as responsible in the particular sense of the word[138] we reserve for *persons*, as opposed to the sense we use to imply that something is part of a causal chain which issued in that for which we are holding them responsible.

But I have shown that neither decisive commitments to lower-order desires, as in Frankfurt's view, nor the ability to act in accordance with one's valuational system, as in Watson's[139], nor the presence of the desire to act in accordance with reasons amongst the motivating factors of one's actions, as required by Velleman, guarantees that it is appropriate to use the former sense of "responsibility" when assessing the actions of such agents. What feature of these selves separates them, then, from creatures or objects which can be held responsible only in the attenuated sense we use when we want to indicate that they are part of some causal chain, and which we would not regard as autonomous at all?

As Wolf points out[140], to continue to insist that an agent such as the victim of a traumatic childhood must be autonomous, in my terms, or responsible, in hers[141], *because* they identify with their motives is to beg the question, because the fact that

---

[138] Wolf, on page 41 of *Freedom within reason*, refers to this sense of the word as "deep responsibility".

[139] 'The contention will be that, in the case of actions that are unfree, the agent is unable to get what he most wants, *or values...*' (1975) 72 *Journal of Philosophy* 206.

[140] *Freedom within reason* 39.

[141] I have already mentioned this terminological difference between my and Wolf's accounts. In this chapter, as above, I shall ignore it, for reasons of ease of exposition, and continue to use the words "autonomy" and "responsibility" more or less interchangeably in what follows.

some people are unwilling to regard such an agent as autonomous is sufficient to cast doubt on the theory as stated at any rate. In the absence of a non question-begging answer from adherents to the subjective view, it seems reasonable to assume that we need to look elsewhere for a theory of autonomy which is able to capture our intuitive sense that identifying with one's motives is not sufficient to confer autonomy on the resulting actions.

In the following chapter, I wish to look at two views which, like the views discussed in the previous chapter, do not require of autonomous agents that they have the capacity to exercise ultimate control over their actions. But the views I examine in this chapter, unlike those of the previous chapter, do not consider agent's own appraisals of their desires as ultimate arbiters in an assessment of whether the agent in question is autonomous or not. The hope is that such views would allow us to explain why agents such as the subservient woman, Son of Sam type cases in which the agent's reflective capacities have been inhibited, and the workaholic should not be regarded as autonomous.

To begin with, I examine Susan Wolf's view that it is the capacity to act in accordance with reason that grants one the status of an autonomous creature.

Wolf's argument for the reason view

Susan Wolf endorses a view of autonomy which she calls the *reason view*. She rejects both the existentialist and the subjective views and argues instead that the capacity which is necessary for autonomous behaviour is the ability to act in accordance with reason. Her view differs from the subjective view of Velleman in that, for Wolf, an agent is autonomous when they act in accordance with what reason *objectively* requires, whereas for Velleman an agent is autonomous when they act on the desire to do what *they consider* reasonable to do.

In order to clarify Wolf's view, let us return to the case of the two would-be rescuers on the beach, which Wolf uses to illustrate the importance of reason in establishing responsibility.[142] In this example both rescuers act in accordance with reason in attempting to save a drowning child, but one of them possesses ultimate control over her actions, and is thus not *bound* to behave in any way. In particular, she is able to reject the reasons which present themselves to her because she is not determined by reason to do anything. The other rescuer, however, lacks the capacity for ultimate control, and is determined to act in accordance with reason. Wolf argues that the ability possessed by the former rescuer 'is an ability that at best seems irrelevant to our status as responsible[143] agents and at worst bespeaks a position directly incompatible with that status'[144]. In doing so she relies on the intuition that it would be odd to hold that the

---

[142]"Is ultimate control necessary for autonomy' section of chapter 1.

[143]"Autonomous" in my terms.

[144]*Freedom within reason* 67.

rescuer who cannot act contrary to reason is less responsible for her actions than the rescuer who can act contrary to reason. On this view, the rescuers' actions are equally praiseworthy in that both acted reasonably, and no further questions about ultimate control need arise.

She rejects then, as do proponents of the subjective view, the idea that ultimate control confers the status of autonomy or responsibility. And, given that she holds that, in contrast to the existentialist view, one *must* be a rational creature to be autonomous, it follows that a creature with ultimate control over its actions could not be autonomous if it lacked the ability to act in accordance with reason.[145] Dogs and psychopaths may, for all we know, possess ultimate control over their actions, but, she argues,[146] this cannot be put forward as a reason for granting them the status of responsible or autonomous beings. So not only is ultimate control not a sufficient condition for the status of autonomy or responsibility, it is not even a necessary condition, for it is conceivable that certain indisputable examples of non-autonomous beings may nevertheless possess ultimate control over their actions.

If we want to know whether we are autonomous or not, according to Wolf, we need to know:

---

[145] 'We might also point out that if one lacks the ability to act in accordance with Reason, one cannot be responsible even if one is autonomous.' *Freedom within reason* 67. Remember that she uses the term "responsibility" where I use "autonomy" and she uses "autonomy" where I use "ultimate control".

[146] *Freedom within reason* 67.

> whether we possess the ability to act in accordance with reason...[T]his amounts to the suggestion that we need to know whether we have the ability to think...well rather than badly...[T]he ability we are concerned with might be described as the ability to do the right thing for the right reasons. The question of whether we have this ability is not so much a metaphysical as a metaethical, and perhaps also an ethical, one. For we cannot answer it unless we know what counts as doing the right thing and having the right reasons...According to the Reason View, having responsible status depends...on a distinctive intellectual power...[147]

By placing the emphasis on the possession of a 'distinctive intellectual power', Wolf takes *competence* of a certain kind to be the mark of the autonomous agent. Competence stands in contrast to the 'distinctive metaphysical power' which is required for autonomy according to adherents of the existentialist view. One might liken Wolf's "competence" condition of autonomy to similar conditions which we require when deciding whether or not someone ought to be held responsible for their actions in contexts other than that of moral responsibility. We might say, for example, that the inexperienced company director who is appointed unwisely to a position of power cannot be held responsible for the company's failure to the same degree than an experienced director could be. This is because the experienced director possesses a greater degree of competence. Likewise, we do not blame an inexperienced sportsperson for his or her poor performance in a pressure situation on the sports field to the same extent that we would blame an experienced player for the same performance. Higher competence brings with it added

---

[147] *Ibid* 70-1.

responsibility. And we cannot make judgements regarding responsibility in these cases unless we know what counts as good business practice or successful sporting tactics.

Wolf regards moral responsibility in the same light as the kind of responsibility discussed in these related examples. If one is to be regarded as morally responsible, one must *know* the moral sphere, and those who judge moral responsibility can only do so if they know the moral sphere. To use her phrase, one must be *normatively competent*. One must be able to understand what morality demands of one in the circumstances, just as the experienced company director or sportsperson knows what is demanded in response to their particular circumstances, and just as judgements of responsibility in the business and sporting contexts cannot be made unless one understands the necessary competences oneself. Along with competence comes responsibility. The kleptomaniac, the subservient woman, certain Son of Sam type cases and the workaholic of the previous chapters can be regarded as lacking in autonomy, not because they lack ultimate control over their choices, but rather because they lack normative competence.[148] They are incapable of doing the right thing for the reasons.

From this it can be seen why Wolf says that responsibility is an ethical, and not a metaphysical, issue. When we praise the sportsperson for the brilliant move on the field of play, we are unconcerned with whether he or she could have done something stupid instead of what they in fact did do. Pointing out that a player *couldnot help* performing the brilliant move in question would hardly serve to show that the player did not deserve

---

[148] I owe this phrase to Gary Watson. Cf (1992) 101 *The philosophical review* 891.

praise. Quite the contrary - we might well regard the player who could not help doing the right thing as *more* praiseworthy than the player who could possibly do the wrong thing under the circumstances. We judge sportspeople by sporting standards, not by whether they are capable of exercising ultimate control over their choices. Likewise, according to Wolf, the moral responsibility of an agent does not rest on the availability of more than one option. Rather, it rests on the availability of one very particular option - the ability to do the right thing for the right reasons.[149] Responsibility is adjudged, on this view, according to ethical standards, and not according to whether one possesses a particular kind of metaphysical power, namely the ability to exercise ultimate control over one's choices.

For Wolf, then, if one does in fact do the right thing for the right reasons then one must be held to be a responsible or autonomous agent. The question of whether one could have done other than what one in fact did do does not necessarily arise for her, as it necessarily does for adherents of the existentialist view. Her view is asymmetrical, however, in that whereas this question does not arise in cases where the agent *does* act in accordance with reason, it does arise in cases where the agent behaves irrationally - in other words in cases where the agent *does not* do the right thing for the right reason. In the latter kind of case the question of blame arises, and it is crucial then to ask whether the agent in question had other options. In this way her view shares an important characteristic with the existentialist view - namely, that both views hold agents to be

---

[149]*Freedom within reason* 68.

blameworthy if they were capable of doing the right thing, but did not do so.[150]

In summary then, Wolf argues that to be held fully responsible, or to be an autonomous agent, one must have been *capable* of acting in accordance with reason, whether or not one in fact did so. If one does in fact act in accordance with reason, one can be praised for one's actions, whether or not one was determined to behave well. On the other hand, someone who fails to act in accordance with reason can only be blamed if they were in fact capable of acting in accordance with reason.

Objections to Wolf's view

Firstly, it should be noted that Wolf is committed to moral objectivity, in that she believes that reason can guide us towards, as she puts it, not only the True but also the Good - that is: she believes that the world contains evaluative as well as nonevaluative facts.[151] Needless to say, this is a controversial position. I don't wish, in this thesis, to enter into the metaethical controversy surrounding the question of whether reason can require certain ethical commitments. But it is important to assess whether Wolf's moral objectivism raises problems for her view. And I want to suggest that it does.

Richard Double has suggested that Wolf's reason view comes close to specifying

---

[150] *Ibid* 68-73.

[151] *Ibid* 72.

the conditions under which one can be regarded as a good person, rather than those under which one can be regarded as autonomous. Although she holds that autonomous agents must not necessarily act virtuously, merely that they be *able* to so act, it's not clear that this is an adequate defence against Double.

He asks us to

> [c]onsider a man who is keenly alive to the question of morality...but usually refuses to give any weight to moral factors in his choices...On the Reason view, is such a person responsible for X-ing? One might think not, because he seems not to be doing the right thing for the right reason.[152]

In other words, he asks us to consider a man who knows what's right and wrong, but doesn't see the fact that something is wrong as a reason for not doing it, and, similarly, doesn't see the fact that something is right as a reason for doing it. Could Wolf regard him as an autonomous agent?

On the reason view, he can only be an autonomous agent if he can do the right thing for the right reasons. However, if one doesn't realize that something's being right is a reason to do it, then it's not clear that one is capable of doing the right thing for the right reason. But if Wolf withholds the status of autonomy from such a person, she seems perilously close to doing so because he is *immoral*, and thereby offering us an account of *virtue* rather than freedom.

---

[152](1992) 101 *Mind* 199.

Wolf is caught in something of a dilemma here. If she believes that the person who recognizes, but is indifferent to, moral reasons, is autonomous, she then faces the same difficulties that subjective views face in that she may be obliged to regard people such as the Son of Sam murderer, or other depraved victims of traumatic childhoods, as autonomous. After all, it is possible that the Son of Sam may have been a person who did what he did in full knowledge of the evil character of his actions.

On the other hand, if she denies that one might appreciate moral reasons without being moved by them then her reason view really does begin to look like a theory of virtuous behaviour rather than a theory of autonomous behaviour. And if this is the case, then the reason view does not fulfil the requirements for a complete theory of autonomy, as the term "autonomy" is not simply a synonym for "virtue". It seems quite possible that someone could act autonomously though immorally.

The asymmetry in Wolf's reason view may well also raise difficult questions about the sense in which those who act badly, but are nevertheless responsible, are *capable* of doing other than they do. She would say that one *couldn't* do the right thing if it was neither psychologically nor physically possible for one to do it. On the other hand, one *could* do it if it were *psychologically* possible, even if one were causally determined in terms of physical laws not to do it. In this latter case Wolf would regard one as fully responsible for one's actions as one would be acting autonomously. Because discussion of freedom and responsibility essentially involve psychological terms, she regards

psychological possibility as crucial to assessments of freedom and responsibility.[153]

The distinction between physical and psychological possibility she draws can be explained in terms of the distinction between physical and psychological explanations of human action. It is possible to explain why people do what they do by using physical terms - ie by explaining their behaviour in terms of the matter they are composed of and the physical laws which this matter obeys. It is also possible to explain why people do what they do by using psychological terms - ie by explaining their behaviour with reference to mental items such as beliefs and desires and concepts such as freedom and responsibility. Psychological explanations may make references to laws which these mental items allegedly obey. One may believe that human action is physically determined, in which case one would believe that it is the inevitable outcome of the interaction between physical particles which obey physical laws. Likewise, one may believe that human action is psychologically determined, in which case one would believe that it is the inevitable outcome of interaction between mental (or psychological) items which obey the laws of psychology, whatever they may be. On the other hand, it is *physically possible* for one to perform a particular action if one is not prevented from doing so by the laws of physics, and it is *psychologically possible* for one to perform a particular action if one is not prevented from doing so by the laws of psychology.

Now Wolf is able to argue that one *could* perform some or other action despite being physically determined not to do so because she holds to a particular understanding

[153] *Freedom within reason* Chapter 5 *et passim*.

of "ability" - that by which one "is able to" or "can" perform some action. On the face of it, this is a surprising conclusion to come to. Physical determinism appears to rule out the possibility of alternative actions. If one's behaviour is determined, ie if one's behaviour is the necessary outcome of the laws which the substance one is composed of obeys, then it would seem that what one in fact does is the only thing one genuinely *could* have done. So, if one does not in fact do the right thing for the right reasons, it seems implausible to suggest that one nevertheless *could* have. And if this inference is correct, it is hard to understand how Wolf could ever hold someone responsible for bad actions. Her theory would allow us to praise people for their good actions, but would forever banish blame - a consequence which would raise the question whether the kind of responsibility being discussed bore any resemblence to our everyday notion of responsibility.[154]

She does not believe, however, that the truth of physical determinism really would forbid us from saying that people could have done other than they in fact did do. She attempts to show this by offering a characterisation of the concept of ability which, she believes, will demonstrate the irrelevance of such determinism to the concept of ability. According to her, ability can be characterised by two claims, one positive and one negative.[155] If one claims that somebody has the ability to do something, one claims, according to her, firstly that the agent in question possesses the capacities or skills

[154] *Ibid* 97.

[155] *Ibid* 101.

necessary for doing whatever it is,[156] and secondly that nothing interferes with the exercise of the necessary capacities or skills. She goes on to argue that determinism does not affect either of these two senses of the concept of ability.[157]

Watson points out, however, that if this is a correct characterisation of ability, then even being psychological determined not to perform a particular action does not mean that one *could not do it*. If ability consists in the possession of the necessary skills or talents for the performance of the action in question, as well as the absence of interference, it seems quite plausible to suggest that one be psychologically determined not to do something, while yet being *able*, in Wolf's sense, to do it. There does not appear to be any signficant difference between physical and psychological determinism in their implications for ability, if Wolf's characterisation of ability is to be taken seriously. And this is a problem for her, as it implies that one could possess the ability to act in accordance with reason even if one was psychologically determined to perform a bad action, just as one could still be said to possess the ability to act in accordance with reason if one was physically determined to perform a bad action.[158]

Why should this possibility concern Wolf? Why can she not simply accept that no kind of determinism, psychological or physical, necessarily takes away one's ability to do

---

[156]It does not seem to me that this is a useful advance in the discussion of the concept of the ability, unless it is held together with Wolf's negative claim - that "ability" implies that nothing interferes with the capacities or skills necessary for performing whatever action is under discussion. This is because it seems to me that one is then simply faced with a new problem in characterising what it is to possess a capacity or skill, and that this characterisation will have to look perilously similar to one's characterisation of ability. I do not wish to enter this debate seriously at this point, however, as I believe that it is beyond the scope of this thesis.

[157]*Freedom within reason* 101 *et seq*.

[158] Watson (1992) 101 *The philosophical review* 892.

something? The problem is that this implies that, despite the fact that an accurate description of the psychological laws governing one's behaviour is incompatible with one's doing the right thing for the right reason, one can still be said to be able to do the right thing for the right reason if one possesses the necessary skills and talents for doing so and if nothing interferes with one's doing so - as would be the case if one was generally able to appreciate moral reasons. If one is psychologically determined to perform bad actions then, on the reason view, one cannot be held responsible, but, the reason view also specifies that one must be held responsible if one is capable of recognizing and appreciating the reasons for and against an action. So being both psychologically determined to act badly and being capable of recognizing and appreciating the reasons for and against the view makes the question of whether one ought to be held responsible for the bad action in question very difficult to answer. It is not at all clear whether one really could have done other than what one did.

Watson also points out that Wolf's requirement of indeterminacy for "bad" actions creates the same problem which the existentialist view faced in the first chapter[159]. In other words, if one can only be held responsible for bad actions if one *could have done otherwise*, then the connection between one's deliberations and either of the two actions which follow becomes mysterious. If the same set of deliberations could give rise to two different actions, it's difficult to see what the connection between the deliberations and the actions are. The actions look just like chance happenings, in which case it's unclear why

---

[159] *Ibid* 893.

we should hold the agent in question *responsible* for their actions. Wolf appears to face the same difficulty the existentialist view faces in explaining how agents who have the ability to choose either way, all antecedent conditions remaining the same, can regard either outcomes of their choices as autonomous, and therefore how they could be held responsible for them.

In summary, then, Susan Wolf's reason view appears to suffer from defects of both the existentialist view and the subjective view. She cannot explain to us how it is that the indeterminacy requirement for bad actions escapes the problem that it faced in the context of the existentialist view. She faces a dilemma in which she must either deny autonomous status to depraved victims of traumatic childhoods, as she must if her view is to be an advance on the subjective view; or endorse what looks suspiciously like a theory of virtue. Given these problems, I wish to turn now, finally, to look at John Christman's theory, in which he attempts to improve on the defects of the subjective view.

Christman's view

I have chosen to conclude this thesis with a discussion of JohnChristman's theory of

autonomy as expounded in his article 'Autonomy and personal history'[160] because I believe that it shares characteristics with both subjective views of autonomy and with Susan Wolf's objective view, while providing what I believe to be a (largely) satisfactory expression of the concept of autonomy itself. In what follows I will offer an outline of Christman's theory, followed by a discussion of certain objections to his views, and explain why I believe his view to avoid the pitfalls faced by the views I have dealt with earlier in the thesis.

He asks us to consider what exactly it is that we feel is problematic about a typical example of heteronomy - his example is the case of a newly-converted cult follower.[161] *Why* do we regard someone who returns from a period spent with a religious cult who then 'mindlessly mouths the credo of the sect, showing few signs of her former self'[162] as lacking in autonomy? If one subscribed to a version of the subjective view, one would pinpoint a lack of identification with her desires or values as the crucial issue. On the other hand, if one accepted Wolf's view, one would argue that if the convert could not act in accordance with reason as a result of her indoctrination then we would be justified in regarding her as lacking in autonomy.

The problem with the answer the subjective view gives is immediately apparent. It may well be the case, if the process of manipulation she underwent was thorough

---

[160](1991) 21 *Canadian journal of philosophy* 1-24.

[161]*Ibid* 10.

[162]*Ibid.*

enough, that the convert *would* identify with her new desires in the requisite way. She could, for example, endorse her lower-order desire to serve the cult leader by means of higher-order volitions, or she could perceive the desire to serve the leader as the natural outcome of her values. If one accepted the subjective view, one would then be obliged to regard her as autonomous. Wolf, on the other hand, appears to be intuitively correct in suggesting that the convert's new desires and beliefs would not be rational ones. But is that why we would regard her as lacking in *autonomy*?

According to Christman, what would be most likely to bother us would be the *manner* in which the new desires and beliefs had been acquired rather than what those beliefs in fact are. We do not necessarily regard adherence to religious dogmas, even bizarre ones, as evidence of a lack of autonomy, unless we think that the agent in question did not participate sufficiently in the process of acquiring those beliefs. Put more simply: we think that certain kinds of processes lead to autonomous desires and beliefs and that other kinds do not. Following from this intuition, Christman suggests that 'the key element of autonomy is...the agent's acceptance or rejection of the process of desire formation or the factors that give rise to this desire formation, rather than the agent's identification with the desire itself'[163] and also 'whether any factors are present during these...evaluations which effectively undercut a person's ability to make these judgements about her past.'[164]

---

[163] *Ibid* 2.

[164] *Ibid* 10.

Note, however, that Christman makes particular reference to the *agent*'s acceptance or rejection of the process of desire formation. On Wolf's reason view the acceptance or rejection of this process by the agent is not important for an assessment of his or her autonomy. An agent may accept the process of desire formation and yet lack autonomy, or, conceivably, reject the process and nevertheless be autonomous.[165] For Christman, on the other hand, autonomy is not simply a matter of objective standards, despite the fact that he does believe that there are standards which are necessary to legitimate the agent's attitude towards the process whereby his or her desires were formed. He shares with subjective views the belief that the agent's own attitude to him or herself is crucial in assessing the agent's autonomy. However, there is a significant difference between Christman's view and subjective views in that for Christman it is not the agent's attitude to their *desires* that confers autonomy. Rather it is the agent's attitude to the *process* that led to them having the desires they do.

This is an important difference. Christman takes the view he does because he is concerned to reject a particular feature of subjective views - namely the fact that on subjective views, the 'determination of autonomy can take place simply by *structural* analysis.'[166] Such views accept that 'a person's desires can be determined to be autonomous or not by taking a "time slice" of the person and asking what her attitude

[165]The latter scenario is possible if, for instance, one believes that one is the victim of uncontrollable forces, whereas in fact it *is* possible for one to act in accordance with reason under one's particular circumstances..

[166](1991) 21 *Canadian journal of philosophy* 9.

would be about the desires she has at the time (or whether they are integrated)'.[167] Subjective views, therefore, do not require investigation of the manner in which the agent's desires are formed, but instead simply ask what his or her attitude to the desires in question is at any given moment. According to Christman, on the other hand, how the agent evaluates the desire in itself may have little to do with the process of desire-formation, and therefore with whether the desire is an autonomous one or not. The virtues of focussing on the process of desire-formation rather than on the attitude an agent has towards his or her desires, he argues, are that it eliminates the need for the condition of identification[168], which, as we saw in chapters two and three, creates such problems for the subjective view, and also that it coincides with the intuition that the way in which someone came to have the beliefs and desires they do is the most important feature of assessments of autonomy.

His principal aim in constructing his theory of autonomy, therefore, is to show what the difference between normal and manipulative processes of desire formation is.[169] The crucial feature of the process of formation of an autonomous desire, he says, is that the agent must have been 'in a position to resist the development of [the] desire and...did not.'[170] He sets out three conditions which he believes will capture this requirement:

---

[167] *Ibid* 9.

[168] *Ibid* 10.

[169] *Ibid.*

[170] *Ibid* 11.

(I) A person P is autonomous relative to some desire D if it is the case that P did not resist the development of D when attending to this process of development, or P *would not have* resisted that development had P attended to the process.

(ii) The lack of resistance to the development of D did not take place (or would not have) under the influence of factors that inhibit self-reflection.

(iii) The self-reflection involved in condition (I) is (minimally) rational and involves no self-deception.[171]

Let us look at these conditions in turn:

The first condition spells out the central claim of Christman's theory - it sets out exactly what it is to approve of the process whereby one came to have the desires one has. As he puts it: 'The motivating idea behind the theory is that autonomy is achieved when an agent is in a position to be aware of the changes and development of her character and of why these changes come about.'[172] If one resisted the development of a particular desire one has, then one can be said to lack autonomy with regard to that particular desire. And if one *would have* resisted the process whereby one acquired a particular desire had one been aware of or attended to that process then one can also be said to lack autonomy with regard to that desire.[173]

---

[171] *Ibid.*

[172] *Ibid.*

[173] Obviously there may be difficulties with ascertaining whether in fact someone would have resisted the formation of a desire had circumstances been different, but neither Christman nor I, at present, are concerned in particular with *how one finds out whether an agent is autonomous or not* - ie with epistemological issues.

Now I have mentioned that Christman believes that his theory is an advance on the theories I refer to as subjective views. In particular he believes that his theory eliminates the problems subjective views face in having to postulate ways in which autonomous agents *identify* with their desires. Looking back at the second and third chapters of this thesis, we can recall that the problem that the issue of identification creates for subjective theories is that it appears to bring about an infinite regress. Subjective views claim that autonomy is conferred by critical reflection on one's desires. But the question of the status of the acts of critical reflection which confer autonomy brings about the regress problem. The acts of reflection, however they are described, must surely be autonomous if they are able to confer autonomy. If this were not so then we would have to regard as autonomous people whose acts of critical reflection had been manipulated in some way. But if we are to specify conditions under which these acts of reflection become autonomous acts of reflection we face the regress problem, as specifying such conditions simply pushes the problem of characterising autonomy one step back. The acts of reflection would have to be endorsed by higher-level acts of reflection, which in turn would have to be so endorsed *ad infinitum*.

It may seem, however, that Christman's view is just as likely to succumb to the regress problem. If specifying that acts of reflection or identification on behalf of the agent confer autonomy, as subjective views do, is what causes the problem, then we can be forgiven for thinking that Christman's view takes us no further. This is because, despite his emphasis on the process of desire formation, he still regards an attitude of the

agent's - in this case an attitude towards the process of desire formation - as the fundamental locus of autonomy. One might well ask what resisting or not resisting the development of a desire consists in. Whatever it is, it must involve attitudes or decisions on the part of the agent, and there appears to be no reason why one shouldn't ask whether the attitude or decision involved in resisting or not resisting the process of desire formation is itself autonomous. After all, it is surely possible for one to be manipulated into approving of the process of desire formation. We would need, therefore, to know whether the agent's approval of the process of desire formation is legitimate. And we would need to be sure of this without falling into the regress that the subjective view falls into.

Christman argues that the regress can be avoided.[174] According to him, '...the claim that all accounts of autonomy that include a condition of self-appraisal are subject to a regress depends on the premise that the only account of the authenticity of the acts of appraisal that comprise autonomy must refer to other preferences of the agent'.[175] The views that were discussed in chapters two and three did share this particular feature - that autonomy was allegedly conferred on the agent's preferences by further preferences of the agent. But, argues Christman, it is not necessary for a theory of autonomy to adopt this feature. As he puts it: 'If the act of appraisal of the processes by which a desire developed in an agent is carried out with sufficient self-awareness and minimal rationality

---

[174](1991) 21 *Canadian journal of philosophy* 18-22.

[175]*Ibid* 18.

then that act of appraisal...is sufficient for the autonomy of the desire.'[176] There is no need to postulate another level of approval which endorses the "lower-order" approval of the process whereby the agent forms his or her desires. Rather, one need simply specify that the approval be a rational and self-aware approval.

In this way it is possible to exclude from the status of autonomy cases whereby an agent was manipulated into approving of the desire formation process. Think, for example, of the recent convert mentioned earlier in this section. Such a convert may well approve of the process whereby she came to acquire the new desires she has. She may, for example, believe that falling under the influence of a (in her eyes) benign cult leader is a legitimate (even virtuous) process of desire formation. And yet we may be certain that the methods used to create in her this belief constitute manipulation of a kind. By introducing an objective feature - namely, that the agent's approval must be rational and self-aware, Christman allows us to explain why such clearly heteronomous agents do not merit the status of autonomy, while avoiding the regress problem that subjective views face.

Christman's second and third conditions of autonomy spell out the objective features of his theory. In particular, they specify that the reflection involved in attending to the process of desire-formation must be rational. But, unlike Wolf, Christman softens the rationality requirement so that autonomy demands only that the self-reflection be

---

[176] *Ibid* 18-9.

minimally rational. By "minimal" rationality he means the kind (or degree) of rationality specified by so-called "internalist" theories of rationality. Internalist views, unlike "externalist" views, claim that it is sufficient for individual rationality if the individual in question not have inconsistent beliefs and desires.[177] Externalist views such as those of Wolf discussed earlier in this chapter set, on the other hand, "objective" standards of rationality. They typically require, for example, that an agent's desires be founded on beliefs for which there is objectively adequate evidence.[178]

Minimal rationality is a much less stringent requirement. It is not necessary for one to be in possession of all the relevant facts for one to be minimally rational. And secondly, as explicated by Christman, it is not even necessary for the minimally rational agent to have internally consistent beliefs. This is because, he argues, 'few of us have examined all our beliefs and preferences and tested them for this standard.'[179] Moreoever, if we did go to this length, few of us would pass the test.[180] Instead, minimal rationality as a condition of autonomy requires that the autonomous agent not be guided by manifestly inconsistent desires and beliefs.[181] The terms "manifestly inconsistent" mean here that the minimally rational autonomous agent will not act on the basis of mistaken inferences or the violation of logical laws with regard to preferences or beliefs

---

[177] *Ibid* 13.

[178] *Ibid* 14.

[179] *Ibid.*

[180] *Ibid.*

[181] *Ibid* 15.

which he or she could easily bring to consciousness.[182]

Christman offers two reasons for rejecting objective standards of rationality. Firstly, on page fifteen of 'Autonomy and personal history'[183] he says that he rejects the externalist rationality condition because he wants to 'develop and defend [a] threshold of normal autonomy'. What concerns him here is the possibility that, if an objective standard of rationality is proposed as a condition of autonomy, the concept of autonomy will be subject to the same vagueness and open-endedness that the concept of rationality is. Given that attributions of rationality to people are reasonably fluid, one runs the risk that attributions of autonomy could be equally fluid. It could then be taken to be the case, he argues, that further gathering of evidence could always lead to greater autonomy, and that we would face difficulties in specifying the point at which normative attributions typically associated with autonomy, such as moral responsibility, would be justifiably made.[184] To avoid this, he believes, we must adopt a minimal rationality condition of autonomy.

It seems to me, however, that autonomy *is* a matter of degree.[185] Christman's concern with the problem of open-endedness is a legitimate one if one is dealing with issues such as legal liability, where one would require as specific as possible a line whereby one could divide people into categories for determining whether they could be

---

[182]*Ibid.*

[183](1991) 21 *Canadian journal of philosophy* 1-24.

[184]*Ibid* 15.

[185]Cf Raz 373, where he he says that '[a]utonomy in both its primary and secondary senses is a matter of degree.'

held responsible in law or not. These issues obviously are important in discussions of autonomy. But the concept of autonomy reaches far beyond those contexts in which a line between autonomous and heteronomous categories of people must be drawn. It would be odd, for example, to call a process of psychotherapy to a halt because one had "crossed the line" from heteronomy to autonomy. In contexts such as this one, it seems quite natural to say that one possesses a certain degree of autonomy which one would like to extend.

The second, and, in my opinion most important, reason Christman puts forward for adopting the minimal rationality condition is the problem that setting an objective rationality condition could mean rejecting agents who are 'acting on well-formed, considered and consistent reasons for action'[186] as heteronomous if they failed to meet external criteria of rationality. This would be too close an identification of freedom or autonomy with rationality which would, he argues, make the property of autonomy diverge from the idea of self-government which serves as its intuitive base.[187] Diverging from the intuitive base of the idea of self-government is not merely of theoretical interest. Holding a view of this sort may also have serious political consequences. To underscore the point, Christman approvingly quotes Isaiah Berlin's attack on the identification of freedom with objective standards of rationality where the latter writes:

---

[186](1991) 21 *Canadian journal of philosophy* 15.

[187]*Ibid* 14. Later on on the same page, Christman says that he wants to defend the minimal rationality requirement for autonomy because he thinks that 'the property of autonomy must not collapse into the property of "reasonable person" where the idea of being self-governing is indistinguishable from the idea of being, simply, smart.' Note the obvious difference between this view and that of Susan Wolf.

> '...[O]nce I take this view...I am in a position to ignore the actual wishes of men or societies, to bully, oppress, torture them in the name...of their "real" selves, in the secure knowledge that whatever is the true goal of man...must be identical with his freedom - the free choice of his "true", albeit often submerged and inarticulate, self.'[188]

Berlin expresses here the intuition that autonomy and freedom are essentially about *self*-rule. The autonomy of a state would be guaranteed by recognition of *its own perception* of its interests, just as individual autonomy or freedom is guaranteed by recognition of the right of individuals to pursue their interests *as they see fit*. Defining autonomy or freedom in terms of an objective standard, without regard to the perceptions of those whose autonomy or freedom is at stake, appears as a misunderstanding of the concept itself. This mistake is avoided, on Christman's view, by restricting the degree of rationality one requires of the autonomous agent to a minimal requirement - that he or she must not act on manifestly inconsistent desires.

It is not clear to me here that Berlin's comments are entirely to the point. It may well be politically dangerous for a ruling elite to believe themselves to be the sole arbiters of rationality, but this does not in itself show that rationality is an inappropriate standard for judging the autonomy of individuals. The question this thesis deals with is *what* one would need to be sure of if one were to make an accurate assessment of someone's

---

[188] *Four essays on liberty* 133. Note here that Berlin's use of the term "real self" differs from that of Susan Wolf, in that Berlin uses it in this context to mean the rational self, where rationality is objectively defined. For Wolf, the real self is identified with one's deepest values, whether or not they are rationally defensible.

autonomy. Whether one could in fact be sure of such things is an epistemological issue best left for further investigation. Nevertheless, I would argue that Christman is correct in wishing to stress the importance of self-government to the concept of autonomy. It seems to me, however, that requiring the the agent's *own* approval of the process of desire formation is necessary for according autonomous status to the desires in question, as spelt out in his first condition of autonomy, is sufficient emphasis on self-government. Suspicion of objective reason, as Wolf points out,[189]

> ...assumes that one's freedom of choice would be compromised if one's choice necessarily followed one's Reason. It assumes that insofar as one's Reason is...decisive in determining one's choice, to that extent the choice is not truly and ultimately one's own. These assumptions reveal an implicit conception of Reason as alien to oneself, as a determining force with which one might in principle be in competition. But, holding fast to the broad and essentially normative use of the word Reason, it is not clear that such a view is intelligible.

It might be also be objected that while certain political dangers may well result from an identification of autonomy with objective criteria of rationality, it is just as dangerous, albeit for different reasons, to require only minimal rationality criteria for autonomy. It may be, for instance, that the substantive rationality requirement is necessary to exclude Son of Sam type agents from the status of autonomy. If we know that the Son of Sam did not

---

[189]*Freedom within reason* 58.

resist the formation of his murderous desires, or that he wouldn't have resisted them had he attended to their formation, we may be, on Christman's theory, obliged to regard him and his ilk as autonomous, provided we can be sure that his lack of resistance to the formation of these desires was not guided by manifestly inconsistent desires and/or beliefs or took place under conditions that inhibited self-reflection.

The important question, then, regards the conditions under which his lack or resistance to the desires took place. If his childhood genuinely was traumatic, it may be possible to show that the process of desire-formation took place under conditions that did inhibit self-reflection. And it may be that we could show that the Son of Sam's approval of the process of desire formation takes place either on the basis of manifestly inconsistent beliefs or desires, or the approval lacks sufficient self-awareness. If any of these possibilities turn out to be true, we would have reason, on Christman's view, to regard the Son of Sam as lacking in autonomy. If it becomes clear that the desires were formed under conditions which did not inhibit self-reflection, or that his approval of them is not based on manifestly inconsistent desires and beliefs and does not lack self-awareness, then we will have to bite the bullet and regard the Son of Sam as an evil, yet autonomous agent - perhaps something like the fictional Dr Hannibal Lecter. This, surely, would be the correct response to the Son of Sam case.

In any case, setting an *objective* rationality requirement is of no particular help in excluding Son of Sam type cases from the status of autonomy. One could just as easily require that minimal rationality excludes certain "immoral" behaviour, and therefore shows

murderers of a particular kind to be lacking in autonomy. At stake in a discussion of whether the Son of Sam could be autonomous or not is not the question of whether the correct rationality requirement for a theory of autonomy is a minimal or objective one, but rather whether rationality requirements of either scope entail moral requirements. Susan Wolf believes that they do, but we have already seen from the discussion earlier in the chapter that such a view results in an excessively moralised view of autonomy.

Christman's theory has the (intellectual) virtue of severing autonomy from moral goodness. In this respect it avoids the dilemma Wolf faces as a result of tying autonomy, rationality and virtue together as closely as she does. Double's case of the man who does not regard the rightness of an action as a reason for performing it, despite his awareness of the difference between right action and wrong action, is not problematic for Christman in the way that it is for Wolf. As in the Son of Sam case, Christman would ask whether the man in Double's example approves of the process whereby he formed the desires he has (desires to disregard moral claims in certain cases, presumably) and whether he is sufficiently self-aware and minimally rational. If the answer is yes, then Christman would regard him as autonomous. And, given autonomy's "intuitive base" as the concept of self-government, it is difficult to see why he should not be regarded as autonomous.

A further reason for adopting a rationality standard as a condition of autonomy,[190]

---

[190]Accounting for our belief that compulsive desires are heteronomous is not a virtue of minimal rationality *per se*. Christman's point is simply that a miminal rationality condition is sufficient to account for the abovementioned belief.

Christman believes, is that it captures what is heteronomous about so-called compulsive desires.[191] Compulsive - ie irresistible - desires are typically taken to be paradigm cases of heteronomous desires. However, he argues, it is possible for compulsive desires to be autonomous. Our objections to compulsive desires, according to him, have more to do with the inconsistency which the majority of them involve.[192] In those rare cases where we experience irresistible desires which are not inconsistent with other desires of ours, we do not regard the irresistible desires as threatening to our autonomy. As an example he mentions the case of the sprinter's response to the starter's pistol. It may be that the sprinter *cannot* control the desire to run at the moment that the starter's pistol goes off. But this is hardly problematic. As Christman points out, this compulsiveness is 'part of a consistent strategy'[193] - one which he believes could conform to the conditions of autonomy. Compulsive desires threaten our autonomy only if they are inconsistent with our other desires - if, for example, they thwart the consistent strategies we have, as would be the case if the sprinter also experienced a compulsive desire *not* to run on hearing the starter's pistol.[194]

There is a further reason, however, to suspect that Christman's view represents no advance on the subjective view of autonomy. If it is a virtue of Christman's theory that

---

[191](1991) 21 *Canadian journal of philosophy* 16.

[192]*Ibid.*

[193]*Ibid.*

[194]Christman's example. *Ibid.*

it avoids Wolf's dilemma with regard to Son of Sam type cases, the same can equally be said of the subjective view. Remember that, on the subjective view, if one identifies with one's desires in the requisite way, one must be regarded as autonomous. Clearly, on this view, the Son of Sam can lay claim to autonomous status if he does so identify with his desires. Yet it is this very feature of subjective views that causes Wolf to reject them. Is Christman's view simply a return to the problem of how to exclude from autonomy certain crucial cases of sick selves who nevertheless identify with their desires?

I would argue that this is not so, and that Christman's theory does in fact offer an advance on the subjective view. It seems to me that the problem subjective views face is not *per se* that such views may have to regard Son of Sam type cases as autonomous agents. As I have mentioned earlier, it surely is the case that such people may well be autonomous. The problem that subjective views face is more specific - it is that they have no reason for excluding Son of Sam type cases from autonomy in many instances *where they are indeed not autonomous*. Christman's view, however, offers us a finer-grained analysis of Son of Sam type cases. It is not obliged to regard such agents as autonomous merely because they identify with their desires, nor is it obliged, as adherents of Wolf's objective view are, to reject them as autonomous agents because they identify with evil desires. Rather, Christman's view offers us the tools for distinguishing those cases where such agents are in fact autonomous from those cases in which they are not. It achieves this by the provisos that the autonomous agent must not only approve of the process whereby he or she came to have the desires they do, but

also that they must do so under conditions of minimal rationality and self-awareness.

To see that these provisos satisfy our intuitions about such cases we simply need to reflect on the conditions under which we would regard such deviant personalities as autonomous, and therefore responsible for their actions, and when we would not. I have indicated above how I believe Christman's view is not obliged, as is Wolf's, to regard Double's man who disregards moral claims as heteronomous. Christman's view also avoids the major problem faced by subjective views, in that, on his view, one is not obliged to regard people we would intuitively regard as sick, and therefore not responsible for their actions, as autonomous - an obligation which the subjective view does appear to face. This is because typically sick, heteronomous, agents, even if they do approve of the process whereby they come to have the desires they have, do not do so under conditions of minimal rationality and self-awareness. In those cases where they *do* do so under conditions of minimal rationality and self-awareness, we *are* inclined to hold them responsible for their actions, and thus to regard them as autonomous agents.

Indeed he goes on to say, on page 23, that

> ...for any desire (no matter how evil, self-sacrificing, or slavish it might be) we can imagine cases where an agent would have *good reason* to have such a desire. Hence we can also imagine that the person is autonomously guided by those good reasons in formulating that desire, and so by that token we can imagine it as autonomously formed, given a fantastic enough situation, then it cannot be the content of the preference *itself* that determines its autonomy. It is always the

origin of desires that matters in judgements about autonomy.[195]

Note that although subservience *is* compatible with autonomy, for Christman, subservience must be chosen under the correct conditions for it to be "autonomous subservience". He mentions the way in which people may constrain themselves in order to achieve certain chosen goals that they might otherwise not be able to achieve without those constraints. The example he uses is of undergoing hypnosis in order to give up smoking.[196] Although hypnosis apparently diminishes one's autonomy, one may have autonomous reasons for choosing to undergo it. Such choices may also be made over an entire lifetime. One might, for example, choose to enter a religious order such as the Society of Jesus which demands a vow of obedience.[197] Despite the fact that such a vow may commit one to a lifetime of having important choices made for one by one's superiors, it may nevertheless represent an autonomous desire for certain goods that the novice (rationally) believes to emanate from such obedience. Of course, it could easily be the case that one makes such choices in a manner which failed to meet the conditions that Christman specifies, in which case they could legitimately be regarded as heteronomous.[198]

---

[195]*Ibid.* Raz makes a similar point on page 371 of *The morality of freedom* when he says that '...the autonomous life is discerned not by what there is in it but by how it came to be.'

[196]*Ibid* 20. Indulgence in sado-masochistic activities may have the same character, in that the participant(s) may voluntarily choose temporary submission or subservience to achieve their desired ends.

[197]I do not mean to imply that devotion to a religious order is the equivalent in all ways of subservience or slavery, merely that it shares the apparent tendency to diminish one's control over one's life and, hence, one's autonomy.

[198]Here I think of someone who is committed to a life of subservience because they have been manipulated into believing that the good life requires such subservience.

****

Conclusion

In conclusion, I would like to emphasise that it seems clear to me that the existentialist view faces problems too serious to warrant further investigation. The view that we do enjoy ultimate control is implausible, and the benefits ultimate control would grant us dubious. On the other hand, the other views I discuss in this thesis, all of which do not regard ultimate control as necessary for autonomy, face different kinds of challenges.

The subjective views fail to explain why it is that we do not accord the status of autonomy to those who do identify with their motivations, but who nevertheless appear to be either sick or lacking in autonomy in some other way. Susan Wolf attempts to remedy this problem by postulating that the capacity to act in accordance with reason is what grants one the status of autonomy, but her view of what reason requires is excessively moralized and either ends up describing the virtuous agent or falls foul of her own objections to the existentialist and subjective views.

In the end it seems to me that Christman's view combines the virtues of the subjective and objective views of autonomy. In emphasising the agent's own appraisal of the process of desire-formation, Christman acknowledges the intuitive importance of

self-government for the concept of autonomy, a characteristic his view shares with subjective views. However, by including amongst the conditions of autonomy the requirement that the autonomous agent's reflection on his or her desires be rational and self-aware, his view avoids the regress problem faced by subjective views. Secondly, his view represents an advance on the subjective view in that obviously sick selves, such as certain Son of Sam type cases, who may, on the subjective view, have to be regarded as autonomous, are excluded from the status of autonomy on Christman's theory if their appraisals of the processes whereby their desires were formed are not sufficiently rational or self-aware.

Christman's view also represents an advance on Susan Wolf's reason view in that it neither identifies autonomy too closely with reason nor too closely with virtue. Autonomy is not synonymous with reason for Christman because of the subjective aspect to his theory - the requirement he makes for autonomy that the agent in question must approve of the process whereby he or she formed their desires. Neither is autonomy synonymous with virtue for him, in that he does not require that the rational agent necessarily be moral. Autonomous evil is possible, on Christman's theory, whereas Wolf faces the possibility that she is obliged to regard someone who understands the difference between right and wrong but does not act on this understanding as lacking in autonomy.

In summary then, this thesis attempts to offer a characterisation of what it is to be an individual autonomous agent. Not surprisingly, it does not cover the entire scope of

the concept of autonomy, in that there is little direct discussion of the political aspect of autonomy. I have also restricted myself, by and large, to a discussion of the conditions of autonomy for particular beliefs and/or desires rather than an attempt to explain what the shape of an autonomous life would look like. I do believe, however, that a characterisation of individual autonomy of this kind is necessary for, and points in the direction of, a discussion of the political dimension of the concept. Any political theory which takes the cultivation of freedom or autonomy as of fundamental political importance must have a clear understanding of what precisely the autonomy which ought to be cultivated in citizens consists in. Without such an understanding a political theory of autonomy will remain committed to vague and confusing ideals. And, given that theories which do place a premium on autonomy, such as liberalism, are currently enormously influential, this would be an unhappy consequence.

Bibliography

Berlin, Isaiah *Four essays on liberty* New York Oxford University Press 1969

Berofsky, Bernard 'Review of Susan Wolf's *Freedom within reason*' (1992) 89 *Journal of Philosophy* 202-8

Christman, John 'Autonomy and personal history' (1991) 21 *Canadian journal of philosophy* 1-24

'Constructing the inner citadel: Recent work on autonomy' (1988) 99 *Ethics* 109-24

Cockburn, David 'Review of Richard Double's *The non-reality of free will* and Susan Wolf's *Freedom within reason*' (1992) 42 *Philosophical Quarterly* 383-8

Dennett, Daniel C *Elbow room: The varieties of free will worth wanting* Oxford Clarendon Press 1984

Double, Richard *The non-reality of free will* New York Oxford University Press 1991

'Review of Susan Wolf's *Freedom within reason*' (1992) 101 *Mind* 198-200

Dworkin Gerald 'Autonomy and behaviour control' (1976) 6 *Hastings centre report* 23-8

'The concept of autonomy' in Haller, R (ed) *Science and ethics* Amsterdam Rodopi 1981

The theory and practice of autonomy Cambridge Cambridge University Press 1988

Elster, Jon *Sour grapes: Studies in the subversion of rationality* Cambridge Cambridge University Press 1983

Ulysses and the sirens Cambridge Cambridge University Press 1979

Feinberg, Joel *The moral limits of the criminal law*

Rights, justice and the bounds of liberty Princeton Princeton University Press 1980

Frankfurt, Harry *The importance of what we care about* Cambridge Cambridge University Press 1988

Jeffrey RC "Preferences among preferences" (1974) 71 *Journal of philosophy* 377-91

Lehrer Keith 'Preferences, conditionals and freedom' in Van Inwagen, Peter (ed) *Time and cause* Dordrecht Reidel 1980 187-201

Levin M *Metaphysics and the mind-body problem* Oxford Oxford University Press 1979

Lindley R *Autonomy* Atlantic Highlands, New Jersey Humanities Press 1986

Madden, Kathleen 'Review of Susan Wolf's *Freedom within reason*' (1992) 45 *Review of Metaphysics* 888-9

Meyerson, Denise 'When are my actions due to me?' (1994) 54 *Analysis* 171-4

Mill, John Stuart 'On the freedom of the will' in Morgenbesser S and Walsh J (eds) *Free will* Englewood Cliffs, New Jersey Prentice-Hall 1962

Nagel, Tom *The view from nowhere* Oxford Oxford University Press 1986

Neely, Wright 'Freedom and desire' (1974) 83 *Philosophical review* 32-54

Penelhum, Terence 'The importance of self-identity' (1971) 68 *Journal of philosophy* 667-78

Raz, Joseph *The morality of freedom* Oxford Clarendon Press 1986

Rosenthal, DM 'Multiple drafts and higher-order thoughts' (1993) 53 *Philosophy and phenomenological research* 911

Strawson, Peter *Individuals* London Methuen 1959

Taylor, Charles *The ethics of authenticity* Cambridge, Massachusetts Harvard University Press 1991

'Responsibility for self' in Rorty, Amelié (ed) *The identities of persons* Berkeley University of California Press 1976

Thalberg, Irving 'Hierarchical analyses of unfree action' in Christman, John (ed) *The inner citadel: Essays on individual autonomy* New York Oxford University Press 1989 123-36.

Van Inwagen, Peter *An essay on free will* Oxford Oxford University Press 1983

'The incompatibility of free will and determinism' in Watson, Gary (ed) *Free will* Oxford Oxford University Press 1982

Velleman, J David 'The guise of the good' (1992) 26 *Nous* 3-26

'What happens when someone acts?' (1992) 101 *Mind* 461-81

Watson, Gary 'Free agency' (1975) 72 *Journal of philosophy* 205-20

'Review of Susan Wolf's *Freedom within reason*'(1992) 101 *The Philosophical Review* 890-3

Wiggins, David *Towards a reasonable libertarianism* in Honderich, Ted (ed) *Essays on freedom of action* London Routledge 1972

Wolf, Susan *Freedom within reason* New York Oxford University Press 1990

'Sanity and the metaphysics of responsibility' in Schoeman Ferdinand (ed) *Responsibility, character and the emotions* New York Cambridge University Press 1987 46-62

Young, Robert 'Autonomy and the inner self' in Christman, John (ed) *The inner citadel: Essays on individual autonomy* New York Oxford University Press 1989 77-90.