

Moral Responsibility for Character

Colin Lumb

Supervisor: Dr. Don Ross

Thesis submitted in partial fulfilment of requirements for the degree of Master of Arts

in the Department of Philosophy, University of Cape Town

September 1997

The University of Cape Town has been given the right to reproduce this thesis in whole or in part. Copyright is held by the author.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Abstract

The aim of my thesis is to fill a gap in virtue ethics. I present an account of moral responsibility that is consistent with the core assumption of virtue ethics, that character and not action is the primary locus of ethical appraisal. Virtue ethics typically does not include a notion of moral responsibility. The reason for this omission is that traditionally attributions of moral responsibility are determined by the causal aetiology of our actions. Because virtue ethics is primarily concerned with our characters and not our actions, virtue ethicists typically assume that causal aetiology is irrelevant to ethics. So in order to fill the gap in virtue ethics I need to show how the core assumption of virtue ethics does not require virtue ethicists to hold that causal aetiology is irrelevant to ethics.

The project of filling this gap in virtue ethics is important because virtue ethics is a popular modern ethical theory and in order to fulfil this function it must say something about moral responsibility. It is a deficit of virtue ethics that it provides us with no basis for judgements about moral responsibility. It restricts the scope of ethics to simply grounding moral assessments of our characters. An ethical theory needs to provide a basis for making practical decisions in legal and political matters.

My thesis is that by shifting the locus of moral responsibility from the notion of action to that of character, virtue ethics can include a notion of moral responsibility. I present an account of the notion of *moral responsibility for character*. It operates as an amendment to virtue ethics and is intended to demonstrate how differences in the causal aetiologies of our characters determine the extent to which we can be attributed moral responsibility for them. In contrast to, for example, Classical Utilitarianism

which holds that being morally responsible is an all-or-nothing affair, my account explains how the differences among the causal aetiologies of our characters create a spectrum along which moral responsibility can be assigned to varying degrees. By focusing on the formation of character and not of action as the basis for attributions of moral responsibility we can make use of the notion of a causal aetiology without being forced to abandon the core assumption of virtue ethics.

My method is a standard philosophical method of critical analysis and synthesis of philosophical literature. My account consists of two conditions that are both necessary and together sufficient for making attributions of moral responsibility for character. The first condition, which derives from Harry Frankfurt's theory of moral responsibility, is that we 'identify' with our characters. The second condition, which incorporates Daniel Dennett's notion of a 'narrative self', is that we are to a significant extent 'morally responsible selves'. My conclusion is that we are morally responsible for our characters only if (1) we are to a significant extent morally responsible selves and (2) we identify with our characters. The extent of the attribution of moral responsibility is determined by the extent to which we identify with our characters.

Contents

Acknowledgements	v
Chapter one – Introduction	1
1. Aim of thesis	1
2. Virtue ethics <i>ethical theory</i>	2
3. Outline of thesis	4
Chapter two – Identification <i>etc.</i>	6
1. Frankfurt's theory	7
2. Conflicts between desires	11
3. Identification and character	14
4. Two problems with identification	18
Chapter three – The Self	26
1. Dennett's narrative self	26
2. Is the narrative self real?	32
3. The unity and disunity of the self	41
Chapter four – The Morally Responsible Self	45
1. Frankfurt's omission	46
2. Possible responses from Frankfurt	51
3. Principal authors	56
4. Remediating Frankfurt's omission	63
5. Conclusion	65
References	67

Acknowledgements

Firstly, I would like to thank my supervisor, Don Ross, for his assistance and encouragement.

Secondly, I would like to thank the Centre for Science Development (HSRC, South Africa) for their financial assistance. It should be noted that opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the Centre for Science Development.

Thirdly, I would like to thank the University of Cape Town for their financial assistance.

Chapter one – Introduction

1. Aim of thesis

The aim of my thesis is to fill a gap in virtue ethics. Traditionally virtue ethics is primarily concerned with the moral assessment of our characters. It typically supposes that we cannot be morally responsible. Virtue ethics allows us to make judgements about whether or not our characters are virtuous. Being virtuous is a matter of possessing certain virtues and not possessing certain vices. There is no question about how we come to possess virtues and vices. Moral assessments of our characters are made independently of their causal aetiologies. Traditionally our moral responsibility is determined by the causal aetiology of our actions. Since virtue ethics entails that the notion of character and not that of action is the primary ethical notion, virtue ethicists assume that they must assert that causal aetiology is irrelevant to ethics. They typically assume that our characters are intrinsic and consequently that we cannot be morally responsible for them (Nussbaum 1986).

Filling this gap in virtue ethics is important, firstly, because virtue ethics has become a popular modern ethical theory and, secondly, because it is important – if not essential - that an ethical theory provides the basis for making attributions of moral responsibility. Many philosophers are turning to virtue ethics as the best alternative to the traditionally dominant Kantian and Utilitarian views. If virtue ethics is going to fulfil this function, then it should include an account of moral responsibility. It is not enough that it simply provides the basis for moral assessments of our characters. It needs to provide the basis for attributions of moral responsibility.

I will present an account of the notion of moral responsibility that is consistent with the defining characteristics of virtue ethics. It operates as an amendment to virtue ethics. The possibility of such an amendment is revealed by changing the focus of which aspect of our causal aetiologies is relevant to morality, from that of the formation our actions to that of the formation of our characters. Virtue ethicists must acknowledge that there can be differences between the causal aetiologies of our characters. These differences provide the basis for an account of moral responsibility that is consistent with virtue ethics. My thesis is that by shifting the locus of moral responsibility from actions to characters virtue ethics can include an account of moral responsibility. I will present an account of the notion of *moral responsibility for character*.

2. Virtue ethics

Virtue ethics is typically characterised by a rejection of the traditionally dominant ethical views in modern moral philosophy - Kantianism and Utilitarianism. Elizabeth Anscombe's call for a virtue ethics view in 'Modern moral philosophy' is conducted in terms of an argument against the primary notions of action-based views. Alasdair MacIntyre's point of departure for his argument for a virtue ethics view, is likewise 'the bankruptcy of modern moral philosophy' (Kruschwitz 1987, 6). Although the criticisms are generally more specific, most of the more recent arguments for virtue ethics also adopt this approach. So arguments for virtue ethics are generally conducted in terms of arguments against the other prevailing modern

ethical views, mainly Kantianism and Utilitarianism.¹ The target of such arguments is usually an overemphasis on the notions of duty, obligation, principle or impartiality.² Although virtue ethics is an alternative to Kantianism and Utilitarianism, this is a negative feature of virtue ethics. It is a feature shared by other modern ethical views, like versions of care and feminist ethics.

Following Gary Watson (1990) and Justin Oakley (1996), I take virtue ethics to be defined by two conditions. Together they constitute the necessary and sufficient conditions of an ethical view counting as a virtue ethics. The first condition is what Watson calls a claim about the *explanatory primacy of character* (Watson 1990, 451). The claim is that the notion of character and not that of action occupies the place of explanatory primacy in ethics. This means that a virtue ethics view takes the notions of morally good and bad character to be primary over the notions of right or wrong, or morally good or bad actions. This is in contrast to, for example, standard Utilitarianism that accords ethical primacy to actions (Hartz 1990, 149). The emphasis in virtue ethics is on various states of mind as ethically crucial (*Ibid.* 145). So if a virtue ethics view goes in for the appraisal of action at all, in other words if it is a 'reductionist' virtue ethics view, then 'action appraisal [will be] derivative from character appraisal' (Watson 1990, 452). On such a view '[a]n action is right if and only if it is what an agent with a virtuous character would do in the circumstances' (Oakley 1996, 129). There is also a 'replacement' version of virtue ethics, like the one Anscombe advocates in 'Modern moral philosophy', that replaces the appraisal of action with the appraisal of character (Watson 1990, 451). Virtue ethics is essentially

¹ For instance, (Stocker 1976).

² For instance, (Nussbaum 1992, 5), (Stocker 1976, 454-6) and (Blum 1994).

in this sense a *character-based* as opposed to an *action-based ethics*. Virtue ethics can be seen to provide a response to the question 'What should I be?' as opposed to the question 'What should I do?' As Beauchamp observes, the difference between virtue ethics and the other dominant modern ethical views is a matter of emphasis (1995).

However there is, in addition, a difference in substance. There is, for instance, a version of Utilitarianism, known as 'Character Utilitarianism', that satisfies this first condition, and includes a notion of virtue, but is clearly not a version of virtue ethics. (Watson 1990, 457). It takes virtue to be of instrumental value, as that which contributes to the general happiness. Hence the need for a further condition. The second condition is that an ethical view includes a theory of virtue (or morally good character) that takes *virtue to be of intrinsic value* (*Ibid.* 455). Therefore virtue must be taken to be the sole or at least primary constituent of what is intrinsically valuable. As Oakley puts it, it must be the case that 'the virtues are irreducibly plural intrinsic goods' (Oakley 1996, 139). Therefore, although it typically denies the possibility, there is nothing essential to virtue ethics that excludes the possibility of an account of moral responsibility. It can do this by shifting the locus of moral responsibility from action to character.

3. Outline of thesis

My account is intended to demonstrate how the causal aetiology of our characters determines our moral responsibility. I will analyse the notion of moral responsibility for character in terms of two conditions. Firstly, I will argue that in order to be attributed moral responsibility for our characters we must *really want*

them. In chapter two I will present an account of what it is to really want our characters. I will critically analyse Harry Frankfurt's theory of moral responsibility (1989a) in order to adapt it to provide a necessary condition for making attributions of moral responsibility for character. Frankfurt does not provide a full account of his notion of identification, so I will flesh it out. Secondly, I will argue that we must play a dominant role in our character-formation. In chapter three and four I will present an account of what kind of self we need to be in order to be attributed moral responsibility. In chapter three I present an account of Daniel Dennett's notion of the 'narrative self' (1991a). I will argue that the self is, as Dennett claims, the gravitational centre of a self-narrative. I will critically analyse this notion in such a way as to present what I take to be an acceptable account of the notion. In chapter four, I will use this notion of a narrative self, as the basis for an account of the notion of a morally responsible self. I will argue that the extent to which we are morally responsible selves is a matter of the extent to which we are the 'principal authors' of our 'self-narratives'. My conclusion is that we are morally responsible for our characters only if (1) we are to a significant extent morally responsible selves and (2) we identify with our characters.

Chapter two - Identification

In this chapter I will present a necessary condition for making attributions of moral responsibility for character. My account is an adaptation of Frankfurt's theory of moral responsibility to the notion of moral responsibility for character. It needs to be noted that Frankfurt does not intend his theory to be put to this use. He does not work out a notion of moral responsibility for character, mainly because he holds the view that ethics is only about 'how to behave' (Frankfurt 1989f, 80). He claims that it is exclusively about 'ordering our relations with other people', in terms of the notions of 'right and wrong'. In terms of Williams' distinction, Frankfurt is of the view that ethics addresses the question 'What should I do?' and not the Socratic question 'How should one live?' or 'What should I be?' (Williams 1985). Frankfurt would say that the content of the second question, which is left out by the first, is properly the topic of a third unnamed branch of inquiry outside of both ethics and epistemology. As far as ethics is concerned, he is only interested in right and wrong actions and consequently, as far as moral responsibility is concerned, he is only interested in a notion of moral responsibility for action. As far as our concern for our characters goes, he does not deny that we might be interested in assessing our characters. But he says that this interest in 'deciding what to do about *ourselves*' and 'what is *important to us*' should be restricted to his third branch of inquiry governed by the question 'what to care about' (Frankfurt 1989f, 80-1). Because I am committed to the central claim of virtue ethics, namely that the notion of character is primary to ethics, I think that this view is incorrect. It takes action to be the primary ethical notion. I am operating within the framework of virtue ethics.

I will demonstrate that Frankfurt's theory of moral responsibility is best interpreted as or at least it lends itself to, including a notion of moral responsibility for character. It is worth noting that Charles Taylor (1976, 281), Ferdinand Schoeman (1988, 9) and Joel Kupperman (1991, 51ff.) interpret Frankfurt's theory as applying to the notion of self or character. But they all seem to be saying that Frankfurt's theory applies to the notion of character only insofar as we need to identify with our characters in order to be morally responsible for our actions. This may be taken to mean that his theory includes a notion of moral responsibility for character. But unlike these accounts, I will be using his theory explicitly to ground a notion of moral responsibility for character, and not as the basis for a notion of moral responsibility for action. So, although Frankfurt may not want to use it for this purpose, I aim to show that his theory of moral responsibility can be adapted to provide the basis for a notion of moral responsibility for character.

1. Frankfurt's theory

In terms of Frankfurt's theory we are morally responsible for things when we 'identify' with them. We do this specifically by identifying with our primary desires for them. So we are morally responsible for things in virtue of our identifying with our desires for them. For example, if I identify with my desire to take my neighbours' supper, and I secretly take their supper because of this identification, then I am morally responsible for taking their supper. In virtue of my identification with my desire to take my neighbours' supper, I am morally responsible for doing so.³ The

³ See Frankfurt (1989d, 51ff.) for one of his examples.

greater the extent to which we identify with things, the more morally responsible we are for them.

Frankfurt's theory requires that we have 'freedom of will' in order to be morally responsible for anything. Frankfurt contrasts freedom of will with 'freedom of action' (Frankfurt 1989c, 19ff.). Freedom of action, on the one hand, is the freedom to do what we want. It is also known as 'freedom of choice', and it is what is mostly discussed under the topic of Free Will. It is usually analysed in terms of the ability to perform alternative actions. We can do what we want when we can choose either to perform an action or not to do so. I have freedom of action when, for example, I can either go for a walk or stay at home. It is understood in terms of the absence of compulsive forces. Compulsive forces are forces that determine our actions for us. Our actions are typically said to be determined by compulsive forces - i.e. unfree – whenever we could not have done otherwise than perform the action we did. Frankfurt's account entails that we are unfree when we act *because* we could not have done otherwise, and not simply whenever we could not have done otherwise.⁴ So freedom of will, on the other hand, is not the freedom to do what we want to do, but the freedom to will what we want to will. We can understand our wills as expressive of that which we want. There is a question about whether or not things like state of affairs in the world are things that we want. This is a question about whether or not we have freedom of action. But there is also a question about whether or not our wills are themselves things that we want. Do our wills constitute that which we really want? This is a question about whether or not we have freedom of will. In this respect

⁴ See for instance Frankfurt's discussion of the unwilling as opposed to the willing addict (1989d, 51ff.)

Frankfurt's notion of identification is intended to conceptualise the range of phenomena surrounding what we 'really want' in, as he puts it, 'the most authentic and perspicuous way' (1989h, 164). He also refers to freedom of will as 'freedom of decision', because, in contrast to freedom of choice, it is the freedom to decide to want what we want. What we are deciding about is our wills and not our actions. The object of a decision, as opposed to the object of a choice, is part of us rather than something outside us (*Ibid.*, 172-4). As opposed to being able to have whatever external object we want, freedom of will means our being able to have the wills we want.

Freedom of will does not entail freedom of action. We can be able to want whatever we want without being able to satisfy these wants. For Frankfurt, being morally responsible for our actions is not a matter of whether or not we could have done otherwise (Frankfurt 1989b, 1ff.). It is a matter of whether or not we really wanted to want to perform an action. For this reason, in terms of our initial example, my taking my neighbours' supper does not have to be the result of my desire to do so, for me to be morally responsible. My being compelled by another person to take my neighbours' supper would not, in itself, exempt me from moral responsibility. I would only be exempt from moral responsibility if I did not identify with my desire to take the supper. I would be morally responsible only in virtue of identifying with this desire. So Frankfurt's notion of freedom of will and consequently his theory of moral responsibility is to be understood in terms of his notion of identification.

For Frankfurt, identifying with some element of our inner life means 'making a decision' (1989h, 172) about that element or making a 'commitment' (1989c, 21) to it. That element then becomes something that we *really want* as opposed to something

for which we simply have a desire or want. This distinction between what we want and what we really want entails a hierarchical model of motivation. This means that we have desires that operate on different levels. Our primary or 'first-order desires' are desires for states of affairs in the world. This is in contrast to our 'higher-order desires', which are desires for desires.⁵ The difference in order is a difference in the object of the desire. First-order desires always have some external state or entity as their object. Second-order desires have certain first-order desires as their objects. Third-order desires have certain second-order desires as their object, and so on. For example, my desire for a bar of chocolate is a first-order desire. My desire for this first-order desire is a second-order desire. Some higher-order desires are also 'higher-order volitions', in virtue of the desire being a desire about which lower-order desires we want to be (part of) our wills (Frankfurt 1989h, 164). Higher-order volitions determine what we really want. The higher the order of the desire that we have for a lower-order desire the more we can be said to really want this desire. If the series of these higher-order desires for a primary desire is, at least logically, infinite, then we can be said to have a 'decisive commitment' to it. If there are no further questions about whether or not we really want a certain primary desire, we can be said to really want this desire. In Susan Wolf's terms, this desire would be partly constitutive of a 'real self' (1990). To say that the series is logically infinite, or that there would be no question about what we really want, is to say that, were we to continue to ask at every higher level 'Do we want to want the desire of the previous level to be our will?', the answer would be an unequivocal 'yes'. To the extent that we have a decisive

⁵ As Dennett (1984, 29) claims, it seems to be this ability to form higher-order desires that makes human activity distinctive. He also says that the exercise of this ability is some of the most important thinking that we do (*Ibid.* 36). At least one reason why the formation of higher-order desires is so important is that it determines whether or not we are morally responsible for things.

commitment to a primary desire we have a 'wholehearted identification' (Frankfurt 1989h, 164ff.) with that desire and we would, consequently, be morally responsible for its effect.

Frankfurt does not make a distinction between moral responsibility and other kinds of responsibility. He concentrates on moral responsibility. But if my desire for a bar of chocolate leads me to rush down to the shop, buy one and eat it, no matter how much I really want this desire for chocolate I will not be in the least bit morally responsible for eating the chocolate. My action simply does not fall into the moral domain. Presumably Frankfurt is operating on the assumption that we are talking about moral matters. Therefore we must assume that we are only morally responsible for things that fall into the moral domain. Presumably, then, if they do not fall into the moral domain we would be responsible, but not morally responsible for them. So, under the circumstances sketched above, although I would clearly not be morally responsible for eating the chocolate, I would nevertheless be responsible for eating it. So we are morally responsible for things to the extent that we have a decisive commitment to our desires for them, assuming that these things or the desires for them fall into the moral domain.

2. Conflicts between desires

Having a decisive commitment entails a certain amount of order and harmony amongst our desires. When we clearly really want a certain desire, this means that this desire is ranked over the other relevant desires in our hierarchies of desire.⁶ It is also

⁶ It should be noted that what is meant by our 'hierarchies of desire' is something different from what

likely to mean that we have no other unintegrated desires, outside our hierarchies of desire, that radically conflict with this desire. Our having a decisive commitment to our desires can be confounded by conflicts between desires. There are two possible kinds of conflicts (Frankfurt 1989e, 66). They are what Frankfurt refers to as 'internal' and 'external' conflicts (1989h, 165).

The first kind of conflict is internal in the sense that it takes place within our hierarchy of desires. The resolution of this kind of conflict is purely a matter of reordering our hierarchies of desire. For example, I may want to both walk and sleep. If we accept for the purposes of this example that the two activities are mutually exclusive – walking entails not sleeping and sleeping entails not walking – then determining what I really want entails endorsing one desire and rejecting the other. Resolving the conflict and consequently having a decisive commitment means having clear higher-order desires for only one of these desires. I need not reject the other desire completely, but merely rank the one more highly. This would entail deciding that of the two desires one takes preference.

As we have already noted, endorsing one desire over another does not mean that the endorsed desire will be more causally efficacious (Frankfurt 1989d, 51ff.). The causal strengths of our desires are determined largely independently, at least in the short term, of which ones we want more. When we endorse one desire over another we simply transform a conflict within our hierarchies of desire into a conflict between the relative causal strengths of our desires. Having a decisive commitment does not require that what we really want corresponds to what is most causally

is meant by the account being a 'hierarchical model of desires'. A hierarchy of desires is a preferential ranking of our desires. A hierarchical model of motivation is a model of motivation that analyses desires in terms of levels ranging from lower to higher-order desires.

efficacious.

The second kind of conflict is external in the sense that it is a conflict between desires that are part of our hierarchy of desires and other unintegrated desires. It is a conflict between desires that we really want and desires that we have neither endorsed nor rejected. The latter kind of desire is probably the kind that unconsciously affects our behaviour. For example, I may have a desire for fame that affects my behaviour in ways of which I am not aware, and comes into conflict with my chief desire to be a hermit. My desire to be a hermit operates as my most preferred desire. When I experience any higher-order desire for my desire to be famous, it is likely to come into conflict with my desire to be a hermit.

As in the case of internal conflicts, external conflicts can be resolved by endorsing one desire and rejecting the other. If we cannot somehow integrate our unintegrated desires into our hierarchy of desires, then we need to reject either of the conflicting desires. We must construe the rejected desire as something we really do not want and consequently as not being part of who we are. This is not to say that by rejecting desires, we are either denying or repressing them. The processes of rejecting and of endorsing desires require a kind of self-awareness that is - by definition - absent in the processes of denial and repression. The former processes can only be conducted by examining our desires and asking ourselves what it is that we really want. We must in some sense be acknowledging the presence and causal efficacy of desires in order to reject them. Repression of desires and (probably more so) their denial, entails that we do not acknowledge the presence, or at least causal efficacy, of these desires. We are repressing or denying the fact that there is any conflict between our desires. It is one of the purposes of psychological therapy to confound the process

of repression or denial. The process of rejecting, or endorsing, a desire would presumably be facilitated by it. When we reject desires, we acknowledge that there has been a conflict. By rejecting these desires, we transform external conflicts between our integrated and unacknowledged desires into conflicts between the relative causal strengths of our endorsed and rejected desires. Our rejected desires might continue to causally effect us, but because they are acknowledged desires, they will not upset the harmony of our hierarchy of desires. What we really want will be clearer.

3. Identification and character

I will now present a proposal for applying the notion of identification to that of character. In terms of Frankfurt's theory, we are morally responsible for our actions insofar as we identify with them, by identifying with our desires for them. This would be to have a commitment to these desires. My proposal is that we are, likewise, morally responsible for our characters insofar as we can be said to identify with them, by identifying with the desires that constitute, produce or maintain them. That is to say that we are morally responsible for our characters to the extent that we identify with either (1) the parts that make them up or (2) the desires that either bring about or ensure the continued existence of parts of our characters.

The second part of this proposal corresponds, in a respect, to the traditional notion of moral responsibility. This notion entails that we are morally responsible for our characters only insofar as we exercise causal control over them, that is, by either originating them, as we do our actions, or maintaining their existence. This is a sense

of moral responsibility for character that entails that we can be attributed only very little moral responsibility for our characters. For, firstly, the existence of our characters is, almost entirely, due to factors beyond our control. We find ourselves with our characters. We do not produce them from scratch. Secondly, the continued existence of our characters is largely due to factors beyond our control. The extent to which we actually maintain our characters is relatively small. Although we can exercise deliberate causal control over them, if we did not, it is likely that our characters would continue to exist unchanged.

We do, however, have a relatively substantial causal capacity for shaping our characters. This allows us, over a period of time, to either maintain or change particular traits. So the most substantial sense in which we can be said to exercise causal control over our characters is in the way in which we shape them either by bringing a new trait about or by promoting an existing one. This kind of control is typically indirect. Its effects usually occur as the result of a series of intermediate causes. They are usually not immediate. For example, I may want to become, or continue to be, generous, but this desire is unlikely to have a direct effect on my character. My desire may provoke me to perform generous actions, but it is unlikely to simply cause me to be generous. So the control we exercise in shaping our characters is most likely to be indirect, in that the effects will, in large part, be due to factors beyond our control. This kind of control is not what is normally meant (in terms of the traditional notion of moral responsibility) by our exercising causal control. It is more likely to fall into the category of the former part of my proposal. It is primarily a matter of identification, that is a matter of really wanting particular traits. The latter part of the proposal is likely to be more easily accepted because of the similarities it

shares with the traditional notion of moral responsibility. But this indirect control indicates the point at which my proposal diverges from the traditional notion. My proposal does not require us to have direct causal control over our characters. We have only to identify with relevant desires. They will not necessarily be our strongest desires (Watson 1989, 112) and they, therefore, do not actually have to be causally efficacious.

The former part of my proposal is that we are morally responsible for our characters to the extent that we identify with their parts - or our traits. Something that Frankfurt (1989e, 62) says about 'the passions in a person's history' may help to clarify this part of my proposal. He says that, although a passion may be external to us in the sense that we did not originate it, by a certain process (which he reveals to be that of identification) 'the passion becomes attached to a moving principal within' us and we are no longer 'passive bystander[s]' with respect to it. It is as if the passion 'had arisen in more integral response to [our] perceptions.' Something similar is true of our character traits. Many of our traits are due to factors external to us, in the sense that they are due to factors beyond our control. But, although we do not originate many of our traits, the process of identification makes them internal to us. When a desire is something that we '*really* want', we are not passive with regard to it (Frankfurt 1989h, 163). Even if we have not originated a desire and are therefore 'not responsible for the fact that it occurs', when we identify ourselves with it we take responsibility for the fact of having the desire - 'it constitutes [part of] what [we] really want' (Frankfurt 1989h, 170). So when we identify even with an existing trait, we can be said to be morally responsible for it.

We normally understand the notion of a trait in terms of the notions of a

disposition, attribute, characteristic, tendency, distinguishing feature, etc. We can understand a trait specifically as an 'enduring characteristic of a person that can serve an explanatory role in accounting for the observed regularities and consistencies in behaviour' (Reber 1995, 807). Williams describes our characters in terms of our 'projects and categorical desires' (1983, 14). So we can think of the parts of our characters (with which we make identifications) as *enduring desires*.⁷ For when we say, for example, that someone is generous, we are saying that she has an enduring desire to give freely. When she gives freely, we can explain her behaviour in terms of this enduring desire. So we are still, as Frankfurt advocates, attributing moral responsibility on the basis of our identifying with our desires. Although in Frankfurt's examples the object of identification is usually a particular passing desire, it does not need to be. There is no reason why the object of our identifications cannot be an enduring desire. The relevant similarity is that they are both desires. In addition, although Frankfurt, on account of his conception of ethics, thinks that there is no need for a notion of moral responsibility for character, he does nothing to exclude the possibility. My proposal, in fact, derives from Frankfurt's discussion of this possibility. He claims that '[t]he question of whether [a] person is responsible for his own *character* has to do with whether he has *taken responsibility for his characteristics*' and not whether he has brought these characteristics about (Frankfurt 1989m, 171). His suggestion, with respect to the notion of responsibility for character, is that it is a matter of

whether the dispositions at issue, regardless of whether their existence is due to the person's own initiative and causal agency or

⁷ This could be, at least part of, what Foot has in mind when she emphasises how 'ethically crucial' 'stable and enduring desires' are (Hartz, 1990).

not, are characteristics with which he identifies and which he thus by his own will incorporates into himself as constitutive of what he is (*Ibid.* 171-2).

Therefore, this adaptation of Frankfurt's theory seems unproblematic; at least insofar as his theory is correct. Because my account of moral responsibility for character is in essence the same as Frankfurt's account of moral responsibility, I need to defend this common ground.

4. Two problems with identification

Problems with Frankfurt's theory usually have to do with his notion of identification. Frankfurt expounds the notion in terms that are broadly acceptable. The terms resound with our intuitions. But he fails to provide a closer analysis of the notion, and it seems to be from here that the problems arise. Frankfurt seems more inclined to acknowledge the absence of a closer analysis of his notion, than to respond to the need for one. He says that '[t]his notion of identification is admittedly a bit mystifying, and I am uncertain how to go about explicating it' (1989d, 54), that '[i]t is difficult to articulate what the act of deciding consists in – to make fully clear just what we do when we perform it' (1989h, 172), and that the way in which he invokes the notions of identification and decisive commitment are, at least at times, 'terribly obscure' (*Ibid.* 167). These comments suggest that we should keep in mind that the task of analysing the notion of identification is not an easy one. I will now consider two possible problems with Frankfurt's notion of identification: first, one that he discusses; and second, one that he does not consider.

The first problem is one that Watson presents (1989).⁸ The problem is that the idea of having a decisive commitment to a desire is inadequate for the purpose of defining the notion of identification. Having a decisive commitment is meant to mark the point at which we say that a full identification takes place. It is at this point that we know what we really want. The problem is one of infinite regress. If a decisive commitment occurs in the form of a higher-order desire for a lower-order desire, how do we know that there are not always further higher-order desires that may indicate a contrary commitment? It is obviously not practical, even if possible, for us consciously to entertain desires higher than the fourth or fifth order. For example, I often desire chocolate and do not for health reasons want this desire. I also have relevant third-order desires, for instance not to have such overly health-conscious desires such as my desire not to desire chocolate. I may further want to have desires, such as my desire not to be so health-conscious, because I want a balance between health and spontaneity. I do not, however, normally have desires at higher levels than these. But it is always logically possible to ask about a higher-order desire, 'Do I really want this desire?' (Taylor 1976, 296). This suggests that we never really have decisive commitments and that Frankfurt's attempts to establish that we do must fail. As Watson puts it, the point at which 'an interminable ascent to higher orders is not going to be permitted ...is arbitrary' (Watson 1989, 118).

Frankfurt's solution to this problem, which I think is correct, is presented in terms of a calculation analogy (1989h, 167-8). He claims that the way in which we come to a decisive commitment is relevantly similar to the way in which we come to a

⁸ Taylor (1976, 276), Fischer and Ravizza (1993, 26) and Thalberg (1989, 130) also present versions of this problem.

solution in performing a mathematical calculation. It is also always logically possible to ask of a mathematical solution, 'Is this answer really correct?' We can always redo the calculation to check our answer. We do not, however, conclude from this that we cannot know when an answer is really correct. There comes a definite point at which it is irrational to continue to ask whether or not it is correct. If there are no conflicting answers and there is no other reason to doubt our relevant mathematical abilities, it is rational to conclude that our answer is correct. Likewise, if we come up with a consistent answer to the question, 'Do I really want this desire?', and there is no other reason to doubt that we are relevantly self-aware, it is rational to conclude that we know what we really want. We can justifiably say that we have a decisive commitment. There is no problem of infinite regress because it would be irrational to enter into one.

Frankfurt also clarifies this aspect of the notion of identification in terms of the notion of a 'resonance effect' (*Ibid.* 169). The resonance effect is the way in which a higher-order desire for a lower-order desire is effectively repeated by each of the theoretically endless series of higher-order desires. The higher-order desire is, then, said to resound throughout this series. Frankfurt explains that a decisive identification is such a resonating higher-order desire. The commitment is made without reservation, and it is therefore pointless to inquire as to whether or not the relevant lower-order desire is what we really want. Nothing would, as things stand, require us to change our decision. In practice, following Gerald Dworkins' suggestion, we can characterise an identification in terms of the highest-order desire that we can entertain (1989, 61).

The second problem has to do with what Frankfurt says about our

identification being a matter of our making some sort of decision. He says that it is 'by making a decision that a person identifies with some element of his psychic life' (1989h, 172). A question that arises is this: if identifying with a desire is a matter of making a decision, then can't we simply not make certain decisions and hereby escape attributions of moral responsibility? This is the position adopted by Albert Camus' protagonist in *The Outsider*. Camus' anti-hero presumes to escape moral responsibility by simply not making any decisions. The objection to Frankfurt's theory is that if we can avoid identifying with things, then it is possible for us to deliberately escape moral responsibility for certain things. This would be highly counter-intuitive. We cannot simply choose whether or not we want to be attributed moral responsibility for things.

In response to this objection I will argue that, in terms of Frankfurt's theory, there is no sense in which we can be said to deliberately escape moral responsibility. We are morally responsible for things to the extent that we identify with them. There are two senses in which we can be said to be avoiding identifying with things. There are two ways in which I can, for example, be said to avoid identifying with my not being a vegetarian – i.e. with my being a meat-eater. I can either be unaware of the need to make a decision about my desire to eat meat, or be aware of this need and not make a decision. If, on the one hand, I am unaware of any need to make a decision about whether to endorse or reject my desire to eat meat, then I will not be morally responsible for being a meat-eater, but I will also not have deliberately avoided making this identification. I can, therefore, not be said to have deliberately escaped an attribution of moral responsibility. This situation of being unaware of the need to make a decision would be the kind of situation that, for example, many Nazis would

have found themselves in during the Second World War. Because they were unaware of the need to make a radical decision about whether or not they should be Nazis, they would not have understood, as we do today, what it really meant to be a Nazi. Consequently they would, to the extent to which they were unaware of the need to make a decision, not be said to have deliberately escaped an attribution of moral responsibility. Slave-owners in the American South in the 1820s could be said to be in a similar situation. Male chauvinists of our parent or grandparent's generation could also be said to be in a similar situation. They were not aware that there was a need to make a moral decision about their being racists or sexists, respectively.

If, on the other hand, I am aware that I need to either endorse or reject my desire to eat meat, then I will be morally responsible. If I do make a decision, then I won't have avoided making an identification, and I will be attributed moral responsibility in accordance with this identification. But if I do not make a decision, then I will most likely be doing so deliberately. Insofar as I identify with my desire not to make a decision either way, I will be morally responsible for my omission. If I neither endorsed nor rejected my desire to eat meat, while I was aware of the need to make a decision, I would be morally responsible for not deciding about my desire to eat meat. This does not amount to my deliberately escaping moral responsibility. This is likely to be the situation in which some of the high-ranking Nazis found themselves. They became aware of what it really meant to be a Nazi, something which included slaughtering thousands of innocent people, and neither rejected nor endorsed their Nazism. We would not say that they deliberately escaped moral responsibility. We would clearly say that they were morally responsible for their omission. They did not make a decision when they knew that they should. We may be inclined to think

that they did escape moral responsibility for being a Nazi. But did they? The punishment at the Nuremberg Trials, which amounted to being punished for being a Nazi, turned on whether or not they knew what Hitler was actually doing. The trials sought to determine things like whether or not they had any knowledge of the death camps whilst still being a Nazi. This amounts to determining whether or not they understood what it really meant to be a Nazi. These people were punished as if they had acknowledged identifying with what it was to really be a Nazi. We can, therefore, infer that they were attributed moral responsibility for being a Nazi. They cannot be said to have escaped moral responsibility. So the second sense in which we can be said to avoid identifying with a desire does not, either, provide a sense in which we deliberately avoid moral responsibility. Incidentally, this is the kind of situation that today's generation faces with regard to racism and sexism. There is a very clear need to make a decision about whether or not to be either racist or sexist. We must either endorse or reject any racist or sexist desires. We do not by refusing to do so escape moral responsibility.

This also seems to, at least partly, describe the situation of Camus' anti-hero in *The Outsider*. For example, if we take him to have been trying to avoid identifying with his desire not to be (or at least play the role of) the compassionate son at his mother's funeral, then we would not say that he has succeeded in escaping moral responsibility, by neither endorsing nor rejecting it. For one thing he is, insofar as he identifies with his desire not to make a decision either way, morally responsible for this omission. This is also not an insignificant attribution of moral responsibility. His attribution would only be diminished by the fact that he does to some extent seem to be unaware of the need to make a decision. He seems to be genuinely bewildered as to

why he either should or should not be compassionate. In this respect his situation resembles the first situation. He would then, to the extent that he could be said to be unaware of the need to make a decision, not be morally responsible.

This example suggests the possibility of a third way in which we can be said to avoid identifying with something. It seems that it is possible that we can in certain situations be neither fully aware nor fully unaware of a particular need to make a decision. This would be likely to be quite a common situation in which we find ourselves with respect to moral questions. We know that we need to make a decision about something, but we repress this knowledge. We may even deliberately repress this knowledge. For example, I may realise that there is a moral question about whether or not to be a meat-eater and consequently know that I must make a decision about my desire to eat meat, but at the same time not really feel like making this decision. As a result I may, either deliberately or not, repress my knowledge. This is, however, not genuinely a third way in which we can be said to avoid making an identification. It simply qualifies, to a lesser or greater extent, as partly the first and partly the second kind of situation. It is important to note that deliberately repressing knowledge of this kind is not an easy thing to do. Many different things are likely to 'remind us'. Also this kind of knowledge does not simply disappear at will. If it happens at all, it is more likely to happen as a result of processes beyond our control. So repressing this kind of knowledge provides no easy way of avoiding making an identification. It does not provide a sense in which we can be said to deliberately escape moral responsibility, because it qualifies, to different degrees, as one of the other situations. If, on the one hand, we fail to repress our knowledge about some moral issue, then we are, to that extent, aware of the need to make a decision and we

will, consequently, be attributed moral responsibility in accordance with what I have said about the second situation. If, on the other hand, we manage, by some extensive process, to deliberately repress our knowledge about an issue, then we will be, to that extent, morally responsible for repressing this knowledge. If our knowledge is repressed as a result of some process beyond our control, then in terms of what I have said about the first situation, we will be, to that extent, exempt from moral responsibility for not making a decision and for our relevant desires. So there remains no sense in which, in terms of Frankfurt's notion of identification, we can be said to deliberately escape moral responsibility. Therefore Frankfurt's notion of identification is an acceptable necessary condition of moral responsibility for character. In chapter four I will deal with a different kind of objection to Frankfurt's theory, and consequently present a further necessary condition.

Chapter three - The self

In this chapter I will present a notion of the self. Because I am operating within a virtue ethics framework and because I am committed to the claim that identifying with our characters is a necessary condition for moral responsibility, there are two constraints on this notion of the self. Firstly, shifting the locus of moral responsibility from action to character means that I cannot appeal to our capacity to perform actions to determine the extent to which we are morally responsible. The extent to which we are morally responsible must be analysed in terms of character-formation. So the notion of the self I present has to allow for an account of how our selves can be, to a lesser or greater extent, morally responsible based upon the causal aetiology of our characters, and not of our actions. Secondly, in accordance with what I have said about the first condition for making attributions of moral responsibility, our selves must be such that they are able to identify with our traits. So the notion of the self that I present must include the idea of a preference structure, or hierarchy of desires, throughout which it is possible for certain higher-order desires to resonate. The self must be unified, at least in the sense that a single preference structure is part of it. I will argue that Dennett's notion of a narrative self satisfies both of these conditions.

1. Dennett's narrative self

Dennett holds that the self is a '*center of narrative gravity*' (1991a, 418). The assertion of a close connection between the notion of the self and that of a narrative

should not be strange to us. Our lives are commonly referred to as our life-stories. As MacIntyre points out, the structure of our lives, starting with birth and ending with death, lends itself to the structure of a narrative with a beginning, middle and end (1981, 191). Our lives are even more closely intertwined with narratives. As Barbara Hardy reminds, us 'we dream in narrative, daydream in narrative, remember, anticipate, hope, despair, believe, doubt, plan, revise, criticise, construct, gossip, learn, hate and love by narrative' (*Ibid.* 197). Telling the story of a death, a break-up or our day enables us to make sense of the event. As Taylor observes, 'we understand ourselves inescapably in narrative' (1994, 57). As we do our lives, we make sense of most things by appeal to the form of a narrative. But Dennett's claim asserts an even closer connection between the self and a narrative. He claims that the self *is* the gravitational centre of a narrative. We are each gravitational centres of that which I will refer to as our *self-narratives*.⁹

But what does it mean to say that the self is a centre of narrative gravity? What does it mean to say that what I am is a centre of narrative gravity? Am I the author of my story, its narrator, or am I its protagonist or central character? The question is whether I actively construct my story, and consequently my self, or whether I am a passive construct in someone else's story? Do I create my self-narrative, or do I just tell it – are the words put in my mouth – or am I the main character, but nevertheless a constructed one in my self-narrative? Dennett seems at times to be saying that we are active in the process of constructing our selves, and at other times to be saying that we are passive in this process. He says that we are authors in the sense that, insofar as our

⁹ I introduce this term for our narratives in order to avoid any confusion with other narratives, specifically autobiographies, which I refer to below.

selves are like spiders' webs, we are 'novice self-spinners' (1991a, 428). But he also describes us as 'theorists' fiction[s]' (*Ibid.* 429). The answer is that each of us is both co-author of, and central character in, our self-narratives. In MacIntyre's terms we are each both 'actor' and 'author' (1981, 198). We are both creators of our self-narratives and the creations of a collaborative venture.

We can understand this in terms of Dennett's spider's web analogy. Dennett says that '[o]ur tales are spun, but for the most part we don't spin them' (1991a, 418). We are to a small extent self-spinners. Each of us, like the spider, does some web spinning. To this extent we create our self-narratives and consequently – because our tales in turn 'spin us' (*Ibid.*), we are their product – we create our selves. By saying things about our selves, about how we do or do not want to be – either in thought or to others – we contribute to our self-narratives, and consequently to the construction of our selves. But for the most part, unlike most spiders, we do not spin alone.¹⁰ We are co-authors of our self-narratives. Other people also say things about us. Other people's conceptions of our selves contribute, either directly or indirectly (by affecting our conception of ourselves), to our self-narratives. Our self-narratives are for the most part the result of interactions between these kinds of social processes and our brains (*Ibid.*, 429). So each of us is to some extent a self-spinner, and for the most part a spun-self – the result of forces beyond us. In addition to being co-authors of our self-narratives, each of us is the central character in our self-narratives.

Is all of this implicit in the notion of a centre of narrative gravity? The question is whether or not this notion of the self entails that each of us is partly author

¹⁰ This is where the analogy breaks down. We may be like spiders insofar as we spin our tales. They spin webs and some may even be assisted by other spiders in the spinning of their webs. But spiders are not themselves 'spun' by their webs, in the way in which Dennett says our tales spin us.

of, and partly character in, our self-narratives. Is this part of what it is to be a gravitational centre of a self-narrative? Thinking of our self-narratives as written autobiographies may be helpful here. What would we say is the centre of narrative gravity of, for instance, Nelson Mandela's autobiography *Long Walk to Freedom*? Well, Nelson Mandela, of course. But do we mean Mandela the author, or Mandela the narrator or Mandela the central character? The answer is all three. The gravitational centre of this narrative consists of all three of these aspects of our experience of the narrative. Reading the narrative, we are aware that Mandela wrote it. We encounter him as authorial presence. He is also telling us the story. We encounter him as narrative presence. We are also aware that the narrative is about Mandela. We encounter him as its central character. The ways in which we encounter Mandela himself in this narrative, are the ways in which we encounter the gravitational centre of the narrative.

Another similarity between this narrative and our self-narratives is that it was a collaborative venture. Mandela had a co-author - Richard Stengal. So, although we might say that Mandela is the principal author of his autobiography, he is not its sole author. Our self-narratives are, likewise, collaborative ventures. Both self-representations and self-interpretations, and others' representations and interpretations of us form them (*Ibid.* 416). This raises the question of the extent of our authorship in our self-narratives. Are we relatively minor contributors, or are we each the principal authors of our self-narratives? On the one hand, as we have already noted, our self-narratives are, for the most part, the result of interactions between certain kinds of social processes and our brains. This would seem to suggest that, as MacIntyre claims, 'we are never more (and sometimes less) than the co-authors of our own narratives'

(1981, 199). On the other hand, there is an obvious sense in which we are each the principal author of our self-narratives. My self-narrative has the property of being my story. As such there is an assumption, on the part of all that encounter my self-narrative, that I am its (sole) author. It is *as if* our self-narratives issue forth from our selves (Dennett 1991a, 418). However, neither of these suggestions capture the sense in which we each may or may not be the principal authors of our self-narratives. This is the topic for chapter four, in which I will use the notion of the principal author of a self-narrative to explain how we can, to a lesser or greater extent, be morally responsible selves.

I would like, at this stage, to raise a point about the extent to which we contribute to our self-narratives. My point is that we are typically not the sole contributors to our self-narratives. Taylor, for instance, seems to overplay the extent to which we contribute to our self-narratives. He seems to be saying that we are essentially the sole authors of our self-narratives. He says that our selves are basically our descriptions of our selves and our self-understandings and self-interpretations (1994, 34). He claims that the self is defined by our 'strongly valued preferences' (*Ibid.* 30). He says that it is in terms of these 'commitments and identifications' that we each determine for ourselves what we really want (*Ibid.* 22). This is correct insofar as this is an aspect of our selves. But our self-narratives are partly constituted by the activities of our preference structures. It is interesting, in this regard, because it is apt, that Frankfurt (1989m, 170) says that by resolving the conflicts between our desires, by 'acts of ordering and of rejection – *integration* and *separation*' – we 'create a self out of the raw materials of inner life.' Although this is partly correct, Taylor and, in this passage, Frankfurt seem to underplay the extent to which others contribute to our

selves. Taylor seems to restrict this social aspect of our self-narratives to the very basic level of a shared language. Other people seem to effect our self-narratives only insofar as they figure in 'webs of interlocution' or as part of a 'defining community' that makes it possible for each of us to form our self-narratives (1994, 36). He says that we have to accept a shared 'language of interpretation' (*Ibid.* 34). But other people's contributions to our self-narratives can be more immediate. They can be co-authors in our self-narratives. It seems obvious that other people's representations and interpretations of us contribute directly to who we are. We have only to think of the way in which we can be said to be genuinely different in different people's company. Other people can play an immediate role in the formation of our self-narratives. There can even be extreme cases where another person, as opposed to the person whose self-narrative it is, is almost the sole contributor. Cases of parents who are extremely manipulative over their children may be of this kind. Perhaps Taylor is acknowledging this point when he says that our commitments and identifications act as a 'frame or horizon' (*Ibid.* 36), presumably within which other peoples' contributions fall. But whether or not he does acknowledge the point, it needs to be noted that other people can contribute to a great extent to our self-narratives.

I would like to return to the analogy between written autobiographies and self-narratives. For there are also relevant differences between these two kinds of narrative. Unlike our self-narratives, written autobiographies are unlikely to be able to facilitate the many sources which contribute to our self-narratives. It is also not possible for us to recall every one of our self-representations and self-interpretations. Even a collaborative autobiography cannot incorporate the contributions of every person with whom we have ever come into contact. But our self-narratives have a

spatio-temporal open-endedness to them. They are, at least while we are alive, works in progress. For each of us our self-narratives are constantly being formed through 'myriads of attributions and interpretations (including self-attributions and -interpretations)' (Dennett 1991a, 426). Each of our self-narratives is part of other 'interlocking narratives' (MacIntyre 1981, 202). This process can even continue after death (Dennett 1991a, 430). The closed form of a biography cannot possibly incorporate all of these elements, while our self-narratives generally do. Our self-narratives are unlike autobiographies in that we are never likely to be able to 'read' our entire self-narratives. When we say things about our selves and each other we are continuously contributing to each other's self-narratives. Our self-narratives are constantly being formed by diverse sources. The only unity that our self-narratives can have is that of our selves, the unity of a single gravitational centre.

Therefore the idea of a centre of narrative gravity entails a rich notion of the self as in each case both author (possibly principal author) of a self-narrative, and its central character. Although there are these different aspects to the narrative self, they amount, at least in normal cases, to a single self. In abnormal cases these different aspects may come apart. For example in cases of brain-washing someone may be the central character, but not an author of his self-narrative. But in normal cases, each of our self-narratives entails a single self who is both author of, and central character in, a self-narrative. A self-narrative encourages us 'to (try to) posit a unified agent whose words they are, [and] about whom they are: in short, to posit a *center of narrative gravity*' (*Ibid.* 418).

2. Is the narrative self real?

The narrative self is bound to strike many of us as too *unreal*. I do not feel like a centre of gravity. If there is anything that feels real to me it is I. It feels as if my self is in some sense inside me. Dennett does seem to be making the point, at least in *Consciousness Explained*, that we are not in the normal sense real. He says that the self is an 'abstraction' (*Ibid.*). I will argue that we are real in the most basic possible sense, as real patterns, and that this is as real as we can possibly expect to be.

In *Consciousness Explained* Dennett's response to the question is to the effect that the narrative self is 'real' in the sense that it is a special kind of abstraction - 'not any old mathematical point' (*Ibid.* 426). He argues that a centre of narrative gravity is 'just as invisible – and just as real' as (for example) the earth's centre of gravity (*Ibid.* 413). We cannot see either, but both provide us with clear explanations for otherwise relatively inexplicable behaviour. In 'Real patterns', Dennett says that selves are 'real because they are (somehow) *good* abstract objects' (1991b, 29). The notion of a real pattern is intended as an account of a good abstract object. A real pattern provides the basis for good predictions and explanations (*Ibid.* 36).

The problem with this line of response is that it maintains the Reichenbachian distinction between *illata* and *abstracta*. Dennett assumes a distinction between concrete and abstract entities. Taylor seems to be making the same assumption when he says that the self 'is not like an object in the usually understood sense', in the way in which 'organisms' or 'hearts' or 'livers' are objects (1994, 34). The problem with this kind of response, then, is that, because of this assumption, no matter how special an abstraction the narrative self is, it is still an abstraction. Therefore the response fails to address the crux of the objection to the notion of a narrative self. We feel that we

are each in the most basic sense real. We exist at the basic level of reality. The problem with maintaining the *illata/abstracta* distinction is that the concrete is usually assumed to constitute the basic level of reality. As Richard Rorty notes, the distinction entails reductionism (1993, 197).

Both Rorty (1993) and Don Ross (unpublished M. S.) address this problem with Dennett's response. They agree that the distinction between concrete and abstract entities needs to be dropped. They are both saying, although in different ways, that Dennett does not take his conception of reality far enough. Rorty says that Dennett simply needs to extend his conception of the self to all entities. He says that Dennett needs to abandon the realism/irrealism spectrum altogether (1993, 198), instead of trying to establish a new kind of 'mild realism' (Dennett 1991b, 30). So Rorty's advice to Dennett is to establish 'the more general claim that *all* objects resemble selves in being centers of *descriptive gravity*' (1993, 189). An object would be, not a concrete particular, but what most of our true beliefs about it are about – a centre of descriptive gravity. Although we would be abandoning talk of the real, our selves would, then, be objects like any other object in the world. However Dennett does not want to follow Rorty's advice, because he is a realist. He holds a distinction between appearance and reality. He believes that there are independent facts out there that are true in themselves and not only for us (Dennett 1993, 234). Even though he thinks that 'when we confront the Observer' this distinction between the 'for me' and the 'in itself' breaks down, Rorty's advice cannot help him to answer the objection, because he wants to maintain the appearance/reality distinction in every other case.

Ross provides a way in which Dennett can both respond to the objection that the narrative self is not real and maintain his realist beliefs. In a sense, Ross makes the

claims that Dennett should have made in 'Real Patterns'. He, like Dennett, wants to 'tie reality to the brute existence of pattern' (1991b, 51). But he does so without the constraints of the concrete-abstract distinction. By maintaining this distinction Dennett is arguing that real patterns also occupy the basic level of reality – along with concrete objects. He is trying to add something non-reductionist to an essentially reductionist metaphysic. His claim that these 'good abstract objects' (*Ibid.* 29) also occupy a place at the basic level of reality runs contrary to the reductionist assumption – which is entailed by his use of the term abstract - that concrete objects occupy every place at this level of reality. He cannot maintain the distinction between concrete and abstract objects and try to argue that certain abstract objects occupy a place at the basic level of reality.

By freeing himself from this distinction, Ross is more able to tie reality to the brute existence of pattern. Ross' claim is that the notion of a real pattern is definitive of what it is to exist: 'To be is to be a real pattern' (unpublished M. S., 9). He, unlike Rorty, does not call for an abandonment of the realism/irrealism spectrum. He takes up Dennett's question about whether the narrative self view is realist or instrumentalist. His answer is that it is, or at least should be, realist. He calls this view 'rainforest realism'; because instead of being committed to a Quinean ontological minimalism - a 'desert ontology' - it is more likely to entail a more densely populated basic level of reality. His advice to Dennett is to define an existent as a real pattern, not a real pattern as a 'good abstract' existent. As real patterns our selves could then be said to occupy the fundamental level of reality.

As Dennett defines it, a real pattern is 'a description of [some] data that is more efficient than the bit map, whether or not anyone can concoct it' (1991b, 34).

For example, my mind is a real pattern, because describing my behaviour in terms of mental states is more efficient than describing it in terms of brain states. My self is a real pattern, because observations about my self are more efficient than the many possible observations about my emotions, sensations and actions as unrelated particulars. A description of my self as, say, extroverted, is going to be more efficient, in that it will provide a better basis for explanations and predictions of my behaviour, than detached observations about my garrulous behaviour or my inability to sit quietly alone for a long period of time. Attributing properties to our selves helps us to understand and make predictions about our behaviour. As real patterns our selves exist at the fundamental level of reality, at the same level that anything at all exists.

The question that remains is how being a real pattern is as real as we can possibly expect to be. There are, at least, two concerns that could be underlying this question. The first is a concern about how the notion of the narrative self is better than standard accounts of the self. The second is a deeper concern about whether or not a real pattern is 'really real'. So, firstly, we may be asking why we can not just stick with something tangible. Do our bodies not come into a notion of the self at all? My body seems to be part of me. Our bodies are also easily identifiable. They are wonderfully perceptible, tangible entities. Are our bodies, or at least our brains, or at least parts of our brains, not somehow part of each of our selves? Or, alternatively, what about souls? Is not each soul a self? The crux of my response to the first kind of concern is that there is nothing to worry about; we are not leaving out anything we shouldn't.

On the one hand, our selves cannot be, not even partly, 'pearls of material substance', 'some spectacularly special group of atoms' in our brains (Dennett 1991a,

430). Hume's observations should go at least some way to dispelling this reductionist hope. We encounter no such selves. Neither by introspection nor by any of our five senses do we encounter a material self in the human body. But this is not to imply that our narrative selves have no relation to our bodies. They are, at least in typical cases, co-existent with our bodies. Our brains produce our selves, in the sense that they give rise to self-representations. Our bodies, and what goes on inside them, also constitute a constraint on the 'myriads of attributions and interpretations' that make up our self-narratives. Not just any attribution and interpretation can contribute to our self-narratives. They must, in a sense, be true. They must correspond to either events in our brains, or events in which our bodies are involved in such a way that the attribution or interpretation makes sense of our behaviour. So our self-narratives are, generally, about events involving our bodies. Details of our self-narratives will correspond to events that happen in or around our bodies. For example, a detail of my self-narrative, like a representation of myself as a farmer, might correspond to certain spatio-temporally located desires I have had about farming. In this way the self corresponds to the spatio-temporal locations of the body. But this does not mean that the narrative self is, or could be seen to be, a material substance. The self is not something that can be material. In Dennett's terms the self cannot be a material thing, because it cannot be encountered from the purely 'physical stance' (1978, 166). It can only be encountered from the 'intentional stance'.

On the other hand, our selves cannot be souls or 'pearls of immaterial substance' (Dennett 1991a, 430). As Dennett observes, it may be natural to infer a soul from a display of wonderful organisation in an organism, but it is a mistake to do so (*Ibid.* 416). Although there is a certain amount of organisation displayed by *Homo*

sapiens, it is insufficient grounds for the inference that we have a soul. Organisation does not mean that there has to be an organiser. Dennett provides the analogy of Eugene Marais' mistaken inference from the organisation he observed in his study of the termite colony, that there exists in each such colony a soul. Incidentally, a little closer to home, Marais made the same inference in claiming the existence of the soul of the ape (1989). Subsequent scientific study has revealed Marais' mistake. We now know that all of this wonderful organisation can be explained in terms of individual genetically programmed organisms doing what comes naturally to them. Without any control centre, but rather as automatons, these organisms are able to behave in amazingly co-ordinated ways by means of simple responses to stimuli. Likewise, we have explanations for the organisation displayed in human beings. Therefore the notion of an immaterial human soul is likewise redundant. There is also a great deal of disorganisation in *Homo sapiens*, which in terms of the structure of the inference to a soul, suggests that there is an absence of an organising soul.

The notion of a soul is also redundant in that it does not help us to explain the degree of unity (and disunity) that the self displays. The notion of an immaterial soul, which persists through time, does not explain how it is that I share a self with myself ten or twenty years ago. If my self twenty years ago had any properties at all, then in what sense is it identical with my self today? They have different properties. The notion of a soul simply adds an extra inexplicable element to any explanation of personal identity. The notion of a soul, also, does not explain how in certain abnormal cases, like that of multiple personality disorder, the self seems to become multiple. What happens to the soul of such a person?

As Dennett argues, even if it is immortality that we are after, the conception of

the self as a pearl of either material or immaterial substance is hopeless (1991a, 430). The notion of a narrative self provides a sense in which the self could be immortal, which is that our self-narratives continue after death. This would clearly seem to be the case with most famous people, who 'live on' in the continued dialogue about them. We have only to consider the new biographies and conversations that continue to arise about someone like Winston Churchill, to grasp the sense in which his self-narrative, and consequently his self, has continued long after his death. For each of us there is the possibility that our self-narratives continue after our deaths in the representations and interpretations of our selves by others. This is the sense in which we each continue in the memory of those we leave behind. As I will argue in the following section, in addition to providing the basis for a hope of immortality, the notion of a narrative self provides a better explanation of the extent to which our selves can, to varying degrees, be either unified or disunified. Thus we can only, and should only want to, expect to be as real as a narrative self is.

My response to the second, deeper concern about whether or not a real pattern is 'really real' is relatively simple. The question is whether or not the notion of a narrative self entails a realist view. This is the question with which Dennett ends 'Real patterns'. My answer is 'yes.' It may seem that, if being a real pattern is a matter of providing a more efficient explanation, then the notion of a narrative self entails an instrumentalist view. For, if there is no restriction on what counts as 'a more efficient explanation', then any information-pattern can count as a real pattern, simply in virtue of constituting a more efficient explanation for one person. For example, if lumping together the orange in my bag, the Great Wall of China and the concept of justice provides me with a more efficient explanation, then this must count as a real pattern.

This is the way Rorty wants to direct Dennett.

The notion of a narrative self does, however, entail a realist view, because there are restrictions on what counts as a more efficient explanation. There is a fact of the matter about what counts as a real pattern. Dennett does not want to reject the appearance/reality distinction (Dennett 1993, 234). Not just any information-pattern can count as a real pattern. Firstly, there are constraints on information flow. There are spatial limitations on information flow. For example, we cannot retrieve information from a black hole. There are also temporal limitations on information flow. For example we cannot retrieve information from before the time of the Big Bang. Secondly, there are constraints on information compression. For example, a series like '0,1,0,1,0,1,0,1,0,1,0,1,0,1...' can be compressed into the information-pattern '(0,1)_i ... (1,0)_n', while a truly random series of numbers is not compressible. It is in this sense that information about our minds is not compressible into information about our brains. So being a real pattern is not simply a matter of providing a more efficient explanation. To be real an information-pattern must in addition, as Ross puts it,

be projectible under at least one physically possible perspective and ... where for at least one of the physically possible perspectives under which the pattern is projectible, there exists an aspect of S [a structure about which the pattern encodes information,] which cannot be tracked unless the encoding is recovered from the perspective in question (unpublished M. S., 9).

These conditions deny instrumentalism and establish Ross's rainforest realism as a variety of realism, because, as Ross explains,

if there is a physically possible perspective from which some

phenomenon recognized by our current working ontology could be more efficiently represented under an alternative ontology, then our current ontology is false (*Ibid.* 10).

As I have discussed the notion, a narrative self is a real pattern in the sense that Ross defines it for Dennett – not in the sense Rorty defines it for him. So the account is realist and not instrumentalist. A narrative self is ‘really real’.

3. The unity and disunity of the self

Finally I want to address the question of the unity of the self. As I indicate above, a notion of the self must allow for explanations of the ways in which we can be unified or disunified. It must be able to explain how the typical self is (more or less) unified. But also, as Dennett notes, the idea that there can only be one self per body is not self-evident (1991a, 422). A notion of the self should, therefore, be able to explain the possibility of more (and less) than one self per body. The adequacy of these explanations is, then, a good test of the adequacy of a notion of the self. The notion of a narrative self passes this test with flying colours.

Dennett rejects talk about selves as containing ‘essences’ and replaces it with talk of ‘similarity clusters’ (*Ibid.* 421). Having a unified self is not, therefore, an all-or-nothing affair. Our selves can be more or less unified. A great deal of disunity would require the postulation of more than one self. But as long as we are in conditions for accumulating a certain amount of ‘narrative richness’ and ‘independence’ we can be said to be ‘fully-fledged’ selves (*Ibid.* 426). This allows for the possibility of disunity, but not too much. A new self would mean a whole new,

separate self-narrative. The practical constraints on this occurring mean that cases of such extreme disunity will be atypical. The one constraint on such a possibility is time. Any candidate would have to get a great deal of 'airplay' before it could be considered an additional self. In normal cases there just is not enough time for more than one self-narrative to unwind. Another constraint is the flexibility of the notion of a similarity cluster. Unlike the notion of an essence it can contain a relatively large amount of dissimilarity before a new similarity cluster has to develop. By its nature, a similarity cluster also cannot develop from isolated exceptions. Our self-narratives can sustain quite a few isolated, strange tales before a new self-narrative is called for. More extended bits of dissimilarity are most likely to be absorbed into the self-narrative. Only in extreme cases will this be impossible. Only in such cases can additional selves be said to develop.

A prevalent test of the adequacy of a notion of the self in the literature is the adequacy of the explanation it offers of cases of split-brain patients. The challenge is that such patients seem to display such disunity in their functioning that it seems necessary to say that there are two selves present – one to each hemisphere of the brain. Dennett's observation is that 'the conditions for accumulating the sort of narrative richness (and independence) that constitutes a[n additional] "fully-fledged" self are not present' (*Ibid.*). There is merely a split in the patient's functioning.

Although the severing of the *corpus callosum* prevents the direct flow of electrical storms from one hemisphere to the other, the two hemispheres are still indirectly connected (*Ibid.* 423-4). Only one self-narrative persists from the time before until the time after the operation. There is no second self-narrative going on 'unseen' in the right hemisphere. The idea of an unseen, or rather unheard, self-narrative is

contradictory.

Another test case is that of multiple personality disorder. The challenge in such cases is that there seems to be more than one self per body. What happens in cases of extreme trauma, usually in childhood and usually of a violent or sexual nature, is that patients seem to develop additional selves in order to face their traumatic experiences. For example, a young boy who is physically abused might develop an aggressive self who ignores the abuse and is able to face it effectively by pretending that it is happening to someone else, while at the same time he might maintain a self who has become a victim. This would be the self that first encounters and endures the abuse. The notion of a narrative self allows for the possibility that genuine additional selves might develop in cases of multiple personality disorder. It would require that a sufficiently rich and independent self-narrative develop. This is presumably something that can only occur in extreme cases of multiple personality disorder.

I also suggest above that a notion of the self must be able to deal with the possibility that there might be less than one self per body. This might occur in the case of someone who is so mindless of themselves and makes so little impression on other people that he or she becomes genuinely 'selfless'. It seems that cases of feral children would be the extreme case of such a kind. The explanation of such a case would be that, because the brain provokes no self-representations or self-interpretations and because no one else has any representations or interpretations of the person, no self-narrative exists and there consequently is no self. But, of course, it would be very difficult for a normal human being to avoid forming any representations or interpretations. Perhaps this is the state of selflessness achieved by Buddhist monks

(Kolm 1981). Although the notion of a narrative self entails that cases of more or less than one self per body will be atypical, if they occur at all, it does allow for the logical possibility of such cases. Therefore the notion of the narrative self explains both how the self is typically unified and how it is possible for it to be disunified.

The notion of a narrative self also satisfies the second constraint, which I set up at the beginning of this chapter, that it includes the idea of a preference structure. The set of self-representations of the form 'I want my self to be X' constitute a preference structure. But most importantly for my purposes, it provides the basis for satisfying the first constraint. It provides the basis for an account of how our selves can be to lesser or greater extents morally responsible selves that is not determined by our ability to perform actions. In chapter four, I will employ the notion of the narrative self to construct such an account of the morally responsible self.

Chapter four – Morally responsible self

We can clearly only attribute moral responsibility for character, or anything else for that matter, to a certain kind of individual. On the one hand, we cannot attribute moral responsibility to animals, infants, young children or extremely insane adults. On the other hand, we usually do attribute moral responsibility to normal adult human beings. We do not have to be looking for a clear line between these cases. They seem to be situated at opposite ends of a continuum of cases of varying degrees of moral responsibility. But the question remains, on what grounds do we say that someone is either to a lesser or greater extent the kind of individual to which we can attribute moral responsibility for his or her character? If we call that which an individual must have in order to be attributed moral responsibility *for* something a *morally responsible self*, then the question: is what is a morally responsible self?¹¹ What selves count as morally responsible selves? What are the grounds for a judgement of 'diminished [moral] responsibility' (Klein 1995, 772)? Firstly, not all human beings are morally responsible selves. Infants and extremely insane adults could not be said to have morally responsible selves. Secondly, not only individual human beings can have morally responsible selves. We cannot assume that humanness *per se* is essential to being a morally responsible self. We must at least leave it open that non-humans can have the necessary properties. Thirdly, as I will argue, simply

¹¹ This is a distinction commonly referred to in the literature. Some instances are: Fischer and Ravizza's (1993, 6) distinction between being a 'morally responsible agent' and 'being morally responsible for something'; Greenspan's same distinction (1988, 81); Dworkin's (1989, 61), which is implied by his notion of 'procedural independence'; Nagel's (1993), which is implicit in the idea of the 'responsible self' as that to which ascriptions of moral responsibility are made; Klein's (1995, 772); Wolf's (1990, 3) distinction between 'responsible beings' and moral responsibility for things, and Fischer's (1988) distinction between being a 'morally responsible creature' and being morally responsible for things.

having a *'selfy agency'* (Dennett 1991a, 416) is also not enough. Not all selves are morally responsible selves. Fourthly, as I will also argue, being able to identify with our characters does not entail the properties necessary for being a morally responsible self. We can identify with our characters and not justifiably be attributed moral responsibility for them. I will argue in this chapter that although identification is a necessary condition for making attributions of moral responsibility for our characters, it is not a sufficient condition. My claim is that having a morally responsible self is also necessary for making any attribution of moral responsibility for our characters. I will present an account of the notion of a morally responsible self, which together with the notion of identification constitutes sufficient condition for making attributions of moral responsibility for character.

1. Frankfurt's omission

The most prevalent objection to Frankfurt's notion of identification as criterion for making attributions of moral responsibility is not that the notion is incoherent, or at least not that it requires closer analysis. It is that the notion does not constitute a sufficient condition of moral responsibility.¹² The objection is that Frankfurt's theory of moral responsibility leaves something out. Richard Double (1991, 33) refers to it as the *'identification problem'*. He says that the problem is that, on Frankfurt's account, there is no way of distinguishing my decisions from decisions that happen "in me". Being able to make this distinction is necessary for attributing

¹² In addition to Wolf (to whom I will refer at length), Christman (1989, 9), Stump (1993, 218-9), Slote (1980, 136-51) and Double (1991, 31ff.) raise this objection.

moral responsibility to someone. In John Fischer and Mark Ravizza's terms, the problem is that Frankfurt only includes a 'freedom-relevant condition' and leaves out a 'cognitive condition' (1993, 8). Frankfurt does not say enough about the background conditions required for moral responsibility. So, he includes a 'specific excusing condition', in the idea of not having identified with something, but he leaves out a 'global excusing condition', in that he does not allow for the possibility that we can be excused from attributions of moral responsibility on the grounds of the kind of selves we are (*Ibid.*, 20). He leaves out the notion of a morally responsible self. I will examine this criticism and consider some responses Frankfurt might make. I will argue (in line with the objection) that Frankfurt's account requires but does not include the notion of a morally responsible self and (in response to the objection) that such a notion can be included along with the notion of identification in a full account of moral responsibility for character.

Wolf raises this objection and considers it at some length (1988). She refers to theories like Frankfurt's, in which category she includes Watson's and Taylor's accounts of moral responsibility, as *deep-self views*. Frankfurt's view is a deep-self view because the notion of identification entails a set of preferences that is constitutive of a (deep) self against which other preferences (shallower selves) are assessed. In terms of the notion of a narrative self, the deep self would be the gravitational centre of a self-narrative. Any shallower selves would not properly be called selves at all. There is almost no chance that single preferences or levels of preference could amount to a sufficiently rich and independent self-narrative to merit the postulation of an additional narrative self. For this reason it might make more sense to refer to these views, as Wolf does elsewhere, as 'real-self views' (1990).

Wolf provides the following example to illustrate her objection to deep self views.

JoJo is the favorite son of Jo the First, an evil and sadistic dictator of a small, undeveloped country. Because of his father's special feelings for the boy, JoJo is given a special education and is allowed to accompany his father and observe his daily routine. In light of this treatment, it is not surprising that little JoJo takes his father as a role model and develops values very much like Dad's. As an adult, he does many of the same sorts of things his father did, including sending people to prison or to death or to torture chambers on the basis of whim (1988, 53-4).

It may be helpful to note that the example is not entirely hypothetical. As Wolf notes, there are actual political leaders whose relationship with their sons, and its consequences, resembles that of Jo and JoJo. The case of Saddam Hussein and his son may serve as such an example. Saddam used to show the young Uday films of prisoners being tortured. Uday seems to have become even more sadistic than his father is, as a result of his upbringing. He is widely known as the most feared person in Iraq. A strikingly similar detail to Wolf's example is that Uday is recently reported to have had the national soccer team beaten for losing a match.

The point of Wolf's example is that, although JoJo values his sadistic character at the level of his deep self, his is not a fully morally responsible self and consequently he cannot be fully morally responsible for his sadistic character. In terms of the example JoJo identifies wholeheartedly with his sadistic character. If we inquired into whether or not he really wants to be sadistic the answer would be a resounding 'yes'. Therefore on Frankfurt's account JoJo would be attributed full

moral responsibility for his sadistic character. But he is 'alienated' from his higher-order desires (Wolf 1993). He is not fully morally responsible for his sadistic character, because he is not to any significant extent the kind of individual to whom moral responsibility can be attributed at all. His identification is, in an important sense, alien to him (Christman 1989, 7). JoJo is not to any significant extent a morally responsible self.

No matter how much we might abhor JoJo's sadistic character we have to accept that he is to a large extent exempt from any moral responsibility for it. Consider the real example of Saddam Hussein's son Uday. He is a despicable character. This is the kind of moral assessment that virtue ethics would advocate. But there is more to be said about Uday. There is a very real sense in which he is not morally responsible for his character and his father is. Uday's upbringing has excluded, if not entirely, then at least to a large extent, factors necessary for the formation of a morally responsible self. If we were to attribute blame for his sadistic character to anyone, it would seem that his father deserves it.

Uday and JoJo are in an important respect like spoiled young children who misbehave. We might quite rightfully call a child a brat for his all-too frequent and unreasonable demands and lack of any signs of appreciation for what he is given. But we would simply be wrong to say that the child is morally responsible for his character. The child is, of course, also unlikely to be able to identify with his character (unlike Uday or JoJo), but it is not for this reason that we would withhold an attribution of moral responsibility for his character to the child. We would withhold such an attribution of moral responsibility for something because as a young child he is not the kind of individual who can be attributed moral responsibility for anything.

He is not to any significant extent a morally responsible self. We would be inclined to attribute moral responsibility for his character to his parents. Similarly we must agree that Saddam Hussein, and not his son, is to a large extent morally responsible for his son's sadistic character.

What is it about JoJo, Uday and the brat, that entails that they are, if not entirely, then at least to a large extent, not morally responsible selves? Wolf's central claim is that deep-self views leave out a 'condition of sanity' (1988, 47).¹³ At this stage I am going to take Wolf's proposal as the criteria for being a morally responsible self. Her proposal is that JoJo is not in a *deep*, slightly unusual sense, sane. It is an unusual sense of the term sane, because we would not normally say that the brat, an infant or the average dog is insane. But, for Wolf, it is because JoJo and these other individuals are not in a certain sense sane and consequently in my terms because they are not, at least to any significant extent, morally responsible selves that they cannot be attributed moral responsibility for anything.

Wolf's definition of her deep sense of sanity stems from the McNaughten Rule of law (*Ibid.* 55).¹⁴ The McNaughten Rule, which has historically been the dominant criterion in legal questions about sanity, defines a sane person as someone who (1) knows what he or she is doing and (2) knows that what he or she is doing is, as the case may be, right or wrong. Wolf identifies a cognitive and a normative component to deep sanity. Accordingly she defines her deep sense of sanity 'as the minimally sufficient ability cognitively and normatively to recognize and appreciate the world for what it is' (*Ibid.* 56). This means that we are sane when we have both 'an accurate

¹³ Wolf goes on to propose a solution for the deep self view, which I will draw on in responding to the objection, in the form of, what she calls, a 'sane deep-self view' (*Ibid.* 56ff.)

¹⁴ Kaufman in (1972, 187) also invokes the McNaughten Rule in this connection.

conception of the world' and an accurate conception of what is valuable in the world. We can summarise these criteria, as Wolf does, as our having an accurate conception of 'the True and the Good' (1990, 75).¹⁵ So JoJo and Uday are not fully morally responsible selves primarily because they have an inaccurate conception of the good. In a broad sense they do not have a real idea of what is good and bad. Their upbringings have determined this. They perceive a sadistic character as being somehow good. They may to some extent also have an inaccurate conception of the world insofar as they perceive the whole world to be like their own, dictator-run countries. The brat seems to be in a similar situation of not knowing what it is to be good and what the world is really like. The difference is primarily one of degree. So Wolf's condition of sanity gives us a conception of what it is that Frankfurt has left out.

2. Possible responses from Frankfurt

I will now consider possible responses that Frankfurt might have to the objection that his theory leaves out a condition of sanity. In other words, I will be examining the possibility that he does implicitly include some such notion in his theory of moral responsibility. I will argue that his theory does not include the requirement that we must be morally responsible selves to be attributed moral responsibility for anything.

¹⁵ This requirement corresponds broadly to what Dennett (1984, 64-5) refers to as the requirement that our beliefs and desires are influenced by 'a clear view of reality and the best of intentions'. For example, we can see that what JoJo primarily lacks is an accurate conception of the good, which corresponds, in Dennett's terms, to being influenced by the best of intentions. It also constitutes 'the yardstick' that, as Taylor (1976, 298) suggests, we are in need of in making 'strong evaluations' of our characters (i.e. making identifications with our characters).

Frankfurt does acknowledge a version of this problem (1989c, 16-7; 1989h, 166) and offer a solution to it (1989c, 21), but, as Double argues, his solution fails. Frankfurt argues that a decisive identification is what distinguishes the person who is morally responsible from the person with wanton second-order volitions. But, for example, JoJo is someone who both decisively identifies with his character and is not morally responsible for it. If the notion of identification does not somehow exempt individuals like JoJo from full moral responsibility for their characters, then the notion is an insufficient condition of moral responsibility.

What we are looking for is a way in which the notion of identification can be said to presuppose the condition of sanity. Frankfurt does suggest that some sort of capacity is required in order to be able to identify with anything. He says that 'it is only in virtue of his rational capacities that a person is capable of becoming critically aware of his own will and of forming volitions of the second-order' (1989c, 17). The possession of this capacity enables him to exempt non-persons, like animals, infants and 'wantons' (1989h, 166), from any attribution of moral responsibility. His notion of a person might correspond to my notion of a morally responsible self. Frankfurt defines a wanton as an individual who has no higher-order volitions. They have only primary desires. In Wolf's terms this would amount to an individual who has no deep or real self. Any inner conflicts are only between their primary desires without their involvement or 'reflexivity' (*Ibid.* 160ff). Reflexivity for Frankfurt is essential for being at all morally responsible. As Dennett puts it, it resides in our ability 'not only to be sensitive to patterns in our environment, but also to patterns in our own reactions to patterns in our environment' (1984, 29). He also describes it as the ability to 'go meta', in that we are able to represent representations, reflect on reflections and react

to reactions. So we are reflexive when we are not simply moved by our primary desires. We have certain higher-order desires about them. We either endorse or reject them. Wantonness can, therefore be seen to be a *minimal* sense of being insane. But JoJo is not a wanton. He is not manipulated by Jo on a 'continuous basis' (Frankfurt 1989e, 53). He has higher-order volitions about his sadistic character. He identifies wholeheartedly with it. So, although being a person may be a necessary condition for being a morally responsible self, it is not a sufficient condition. JoJo can and does identify with his sadistic character. He is therefore in Frankfurt's sense a person. However, as we have noted, he is not a fully morally responsible self.

A second possibility involves a *fuller* sense of being insane. That is the kind of insanity for which we may go to therapists – varying degrees of disharmony in our preference structures. When we have conflicts between our same level higher-order desires, then we can be said to be to some extent insane. The extent of such inner conflicts can vary from something about which we might casually say is driving us insane, like trying without success to make a decision about ourselves, to something for which a psychologist would commit us to a mental institution. Anything that prevents us from making wholehearted identifications exempts us, to some extent, from attributions of moral responsibility. We are all at times likely to be, in this sense, to some extent insane. It is seldom that any of us make wholehearted identifications. But again this is not Wolf's deep sense of insanity. This fuller sense of insanity has to do with how much moral responsibility is attributed to us. The less wholehearted our identification with something the less the attribution of moral responsibility for that thing. JoJo could not be said to be, at all, insane in this fuller sense. He identifies wholeheartedly with his sadistic character.

There is a third possibility that, although unrelated to the notion of identification, may be seen to establish that Frankfurt does implicitly hold that being a morally responsible self is a necessary condition for an attribution of moral responsibility for something. The proposal is that Frankfurt's notion of 'rationality' amounts to Wolf's condition of sanity.¹⁶ If this were so, then, in order to overcome the objection that he leaves this condition out, we could simply incorporate Frankfurt's notion of rationality into his theory of moral responsibility. Frankfurt says, in a slightly different context, that part of our being rational individuals is that certain things are 'unthinkable' for us (1989i, 186ff.). This entails that as rational individuals, or at least insofar as we are rational, we will have a certain kind of self: We will not be able to be or do certain things. The intuition behind the notion of the unthinkable is evoked by, for example, the sense in which we could not cut a newborn baby's throat. As rational individuals there are certain possibilities that are excluded, for us, from serious consideration. Applied to the notion of identification this would mean that there would be limits on the kinds of things with which we could identify. We would not be able to bring ourselves to endorse certain things (Frankfurt 1989n, 187). Certain things would simply be beyond the bounds of what we could really want. So, presumably, the kind of example that Frankfurt might have in mind would entail that we could not really want to be a sadistic dictator like Stalin.

However the problem with this proposal is that it implies that there are certain kinds of things, presumably mostly bad things, for which we simply could not be morally responsible. This is counterintuitive. Surely it is at least logically possible for

¹⁶ This is a different, fuller sense of rationality to that discussed under the first possibility – the minimal sense of a rational capacity in terms of which Frankfurt characterises a person.

us to be morally responsible for anything. As Double (1991, 67) suggests, we cannot include the requirement, as a precondition for attributions of moral responsibility for things, that the subject is a 'moral robot'. This is, as Watson suggests, paradoxical (1993, 131). It cannot be the case that being a morally responsible self and making an identification entails not having a morally bad character. We would not want to say that, for example, Stalin could not be morally responsible for his character simply because it was a particularly bad character. We believe that, even if Stalin himself was not rational, it is possible to have both a sadistic character and a morally responsible self. We could, as morally responsible selves, perhaps as an intellectual exercise, rationally cultivate a sadistic character. Neither the McNaughten Rule nor Wolf's condition of sanity implies that it is impossible for us as sane individuals to think certain things. We are inclined to want to believe that there is something deeply wrong with an individual who can, for example, over an extended period, premeditatively murder many innocent people. In short, we are inclined to infer that an individual with an extremely bad character must be insane. They must have suffered some sort of abuse themselves. But we also have to accept that this is not necessarily the case. We are inclined to hope that it is, but we must accept that it is possible that they are not insane. There is no reason why it would be impossible for such an individual to be as sane as any normal adult human being and do these (unthinkable) things. So although Frankfurt's notion of rationality may include Wolf's notion of sanity, it includes more than what we want in a notion of a morally responsible self. Therefore as it stands, Frankfurt's theory of moral responsibility does leave something out. We cannot use only Frankfurt's terms to fill this gap. We need to introduce something like Wolf's condition of sanity.

3. Principal authors

Up until this point we have understood the notion of a morally responsible self exclusively in terms of Wolf's notion of sanity. It seems that we should simply conclude that the identification condition and Wolf's sanity condition can together form a complete account of moral responsibility for character. We would then have something like what Wolf calls a 'sane deep-self view'. This is in large part what I am doing, but I am going to articulate the notion of a morally responsible self in terms other than Wolf's. Firstly because, although Wolf does set up a sane deep-self view in 'Sanity and the metaphysics of responsibility', she ultimately does not hold such a view. Her arguments are ultimately aimed at establishing the 'reason view' she puts forward in *Freedom within Reason*.¹⁷ Although I do not think that using her terms will make my account incoherent they are formulated with another view in mind.¹⁸ Secondly, I will formulate the notion of a morally responsible self in the terms I have used in chapter three, because this will facilitate a more unified account of moral responsibility for character. I will articulate the notion of a morally responsible self in terms of the notion of a *principal author* of a self-narrative.¹⁹ I will argue that a morally responsible self is a self who is principal author of her own self-narrative. I will relate my discussion to Wolf's example of JoJo, but I will begin by introducing a

¹⁷ This view is also known as 'the reason-responsiveness' view, which Fischer (1988) also holds.

¹⁸ Frankfurt would probably reject this reason view on the grounds that acting for good reasons cannot be the criterion for attributing moral responsibility (advocates of this view claim that it is), because even a wanton can reason and consequently act for good reasons (1989h, 170). I am inclined to think that the two views are not mutually exclusive, but that Frankfurt's view operates at a more basic level than Wolf's. Frankfurt does suggest that reasons operate at a more superficial level than the will (1989i, 189). I would go along with this suggestion.

¹⁹ I am indebted to Don Ross for suggesting the term 'principal author' and for helping me to develop it.

new, and real, example in terms of which to conduct my discussion.

The example is the case of Robert Harris (Fischer 1993, 1-4 and Watson 1993, 131-7). Harris was convicted of murder, sentenced to death and duly executed. His crime was the brutal murder of two 16-year-old boys without any provocation. He and his brother intended to use the boys' car for a bank robbery. Harris jumped into their car in a parking lot, pointed a gun at the driver's head and ordered him to drive out of town. Harris' brother followed in their car. They stopped at the roadside, Harris explained his intention and assured the boys that he would not harm them. They had only to walk back into the nearby town. As the boys turned to do so, Harris shot the one boy in the back of the head, pursued the other, killed him and returned to kill the first boy. He had no reason to kill the boys. He was consistently unrepentant, even jovial, after killing the boys. Both professionals involved in the case and Harris' fellow inmates on death row were emphatic and unanimous in their opinion that, if there ever was someone who deserved the death penalty Robert Harris did. His story invoked moral disgust in almost everyone who heard it. He consistently showed no sign of remorse. Even by the standard of his fellow inmates, he was an extreme case. This suggests that he identified wholeheartedly with his murderous impulses. This inference is further supported by his cool responses to other horrific deeds, which progressed in severity, from the point of, as a boy, torturing his pets to death. We can, therefore, assume that around the time when he killed the boys he identified wholeheartedly with his cruel and sadistic character. So, according to the proposed criterion, it would seem that we would have to say that Harris was fully morally responsible for his cruel, sadistic character.

The details of his upbringing, however, have the effect of undermining this

attribution of moral responsibility at a deeper level. They suggest that he was not to any significant extent a morally responsible self. Even the moment of his birth was abnormally traumatic. His father brought about his premature birth by kicking his mother repeatedly in the stomach. His father's provocation was the false belief that the infant Robert was not his son. The belief determined all of Robert's father's cruel treatment of his son. His mother also became progressively more cruel to her son as she came to see him as the cause of her husband's cruelty to her. One of his sisters explained, in an interview, how Robert slowly changed from being the most sensitive of the ten children to being a sadistic torturer of animals and finally a cold-blooded murderer. In Wolf's terms, Harris's sanity was reduced by his upbringing. He had an extremely distorted conception of the true and the good. We can infer that he experienced almost exclusively, pain, injustice, cruelty, abuse and the absence of love.

Now I will examine the sense in which Harris is not a morally responsible self in terms relating to the notion of a narrative self. My claim is that Harris is not the principal author of his self-narrative, because his self-narrative has been formed for him by a few dominant co-authors. These co-authors' contributions to Harris's self-narrative are in important respects the same. Because their contribution was effectively the same, because it was of such a persistent nature and because Harris's self-narrative lacked any contributions contrary to these, Harris never made any substantial contribution to his self-narrative. I will examine particularly the contributions of two co-authors – his father and mother.

The predominant contribution of both of Robert's parents was a representation of Robert as worthless, or more accurately as of negative value. They did not want Robert and thought of him only as a cause of unhappiness. For his father Robert was a

constant reminder of his own worthlessness, because he was a reminder that his wife did not need him. As I have said, Robert's father imagined that Robert was not his son. As a result Robert's father abused him. He cruelly punished Robert for no reason. He would beat Robert apparently on a whim. It is likely that this corresponded to his being reminded of his own worthlessness. For his mother Robert came to be seen as the cause of his father's cruelty towards her. Robert's father beat his mother. The effect on her would have been, amongst others, a feeling of worthlessness. Her response was to ignore Robert, so as not to provoke her husband. Any show of love for Robert would have been perceived, by her husband, to be a show of love for her supposed lover. Her representation of Robert would have been of a worthless, troublesome self. She interpreted him as the cause of her husband's cruelty towards her. Robert's father's interpretation of Robert was as a reminder of his wife's (imagined) infidelity and consequently of her not needing him. These representations and interpretations were of such a persistent and consistent nature that they determined what Robert was.

In terms of Dennett's notion of a narrative self our brains give rise to self-representations. We naturally represent our selves in presenting our selves to ourselves and to others, and in doing so we try to form a coherent self-narrative. But, if the people with whom we interact consistently and persistently represent us in a certain way, then we do not have to do much self-representing. We simply represent ourselves as they do. We hereby become minor authors in our self-narratives. This is the situation with Robert Harris. The people with whom he interacted most (and it seems that he did not have any other major influences that were contrary to that of his parents) represented him as a worthless and bad (because he was equated with certain

bad effects) individual. Thus Robert's parents were the principal authors of his self-narrative. Because he was not the principal author of his self-narrative he was not, to any significant extent, a morally responsible self.

For almost all of us our parents or parent-figures are the principal authors of our self-narratives early in our lives. But for most of us, relatively early on in our lives, we take over this position. For most of us, our teenage years will be the clearest signs of our taking over this position. Teenagers are renowned for saying things to their parents like "It's my life – leave me alone" or "That's what you want me to be – it's not who I am." So most of us take over the position of principal author of our self-narratives from our parents.

This taking over of principal authorship is evident in the impulse to either 'go off on one's own' or 'be with other people'. The first is an impulse to take over, simply and straightforwardly, the position of principal author. I want to shape my self by myself. The second is an impulse to find my self in others. I take over the position of principal author, almost paradoxically, by allowing in many different co-authors. These co-authors each, mostly for short periods, contribute to my self-narrative by their representations and interpretations of me. I maintain my role of principal author because no one co-author's contribution is sustained enough and similar enough to another co-author's contribution to shape my self-narrative for me. So I am to some extent principal author in the case that I am either the sole contributor to my self-narrative - like the hermit who at an early age shuns society to be alone - or the persistent contributor to my self-narrative amongst a series of other contributors, like the world-traveller who is continuously on the move and meeting new people. The traveller's quest is to find himself. Therefore cases like those of the hermit and the

traveller define extreme cases of principal or self-authorship and consequently of minimally morally responsible selves.²⁰ Along a continuum away from each of these extremes towards the centre are cases of increasing self-authorship and therefore of more morally responsible selves.

There are two factors involved in being the principal author of a self-narrative: (1) being the *persistent author* and (2) having a *set of co-authors*. When we are both the most persistent authors of our self-narratives and we have a large and coherent set of co-authors, then we are to the greatest extent the principal authors of our self-narratives. The cases of only one factor without the other are probably practical impossibilities, but as logical possibilities they function as limiting cases. The closer we get to them the less the extent to which we are principal authors. On the one hand, the case of someone who is the most persistent author of her self-narrative and has no co-authors defines the pole on the hermit-case side of the continuum. It is a practical impossibility because anyone without any co-authors – i. e. someone who had never had any contact with others – would not produce any self-representations. Our self-representations are the result of an attempt to form a coherent self-narrative out of others' representations of us. The hermit is a less extreme case and consequently to some extent principal author of his self-narrative because he has had co-authors – before leaving society. On the other hand, the case of someone who has only a large set of co-authors and makes no contribution herself defines the pole on the traveller case end of the continuum. It is a practical impossibility because having co-authors entails an attempt to form a coherent self-narrative. Contributions by co-authors

²⁰ The notion of principal or self-authorship can be seen to correspond, at least initially, to the notion of self-government, which Christman (1989, 4ff.) suggests is a precondition for attributions of moral responsibility for things.

necessarily give rise to self-representations. The traveller is a less extreme case and consequently to some extent principal author of his self-narrative because he makes some contribution to his self-narrative in response to the contributions of others. So the two poles constitute the limiting cases on the continuum. As we move towards the centre from either extreme, we find cases of increased principal authorship.

The case of Robert Harris lies somewhere closer to the traveller case than to that of the hermit. Harris lacks the persistent author factor to a significant extent. His contribution is undermined by his parents' contributions. He is like the traveller in that his self-narrative is to a great extent determined by the contributions of others. He seems not to have come into contact with a very wide range of responses to himself. Although he may actually have come into contact with many people, their contributions to his self-narrative would have been unlikely to differ much. He was reported to have been antagonistic to all with whom he came into contact. He would have been consistently represented as a worthless and troublesome self by others. Therefore he was not principal author of his self-narrative and consequently he cannot be attributed moral responsibility for his cruel, sadistic character. We are inclined to blame his parents, and the basis for this intuition is that they were the major co-authors of his self-narrative.

The notion of a principal author also accounts for our attributions of moral responsibility in other cases. Firstly, we can see how the notion of a principal author grounds the judgement that JoJo, the sadistic dictator, is not a morally responsible self. His father was in a sense the principal author of his self-narrative. He was the dominant co-author of, and most persistent contributor to, JoJo's self-narrative. Secondly, it explains why someone who spends most of their time by themselves

might not be a morally responsible self. They may not have any significant co-authors. They can literally make up their self-narratives. Thirdly, it explains why infants and extremely insane individuals are not morally responsible selves. They make no contribution to their self-narratives and consequently are not the principal authors of their self-narratives. Fourthly, it explains why animals cannot be morally responsible selves. They cannot be principal authors of self-narratives because as non-selfy agents they do not have self-narratives. They cannot represent themselves at all. Finally, it explains why selves whose characters and choices are determined by a controller of some sort are not morally responsible selves and consequently cannot be attributed moral responsibility for things. The controllers are the persistent authors of their self-narratives. Cases of, what Dennett (1994a, 7-10) calls, 'incompatibilist bugbears', like those of the 'Nefarious Neurosurgen', the 'Cosmic Child Whose Dolls We Are' and the 'Peremptory Puppeteer', would be such cases. I conclude that we are morally responsible selves when we are the principal authors of our self-narratives.

4. Remediating Frankfurt's omission

We can overcome the objection that the notion of identification is insufficient as an account of moral responsibility for our characters, by incorporating into our account the notion of a morally responsible self. My proposal is that we do this in terms of the notion of the principal author of a self-narrative. This means that we are morally responsible for our characters to the extent that we are morally responsible selves who identify with our characters. When we are to a significant extent morally responsible selves, we are to the extent that we identify with our characters, or any of

our traits, morally responsible for them. There is no paradox in the idea that we must be, to some extent, morally responsible selves before we can be attributed moral responsibility for things. Watson (1993, 131) suggests that there may be a paradox in implementing such a condition of moral responsibility. Analysing the condition in terms of the notions of rationality and the unthinkable would result in a paradox because the condition would imply that we could not be morally responsible for certain, mostly bad, things. But the idea of principal authorship does not exclude the possibility that we can both have extremely evil characters and be attributed moral responsibility for them. It may, however, be a happy coincidence that many extremely evil characters are also accompanied by non-morally responsible selves. For example, it happens that selves with characters like Robert Harris' are not their own principal authors, but there is no reason why they cannot be.

An objection to the incorporation of this notion of a morally responsible self into my account of moral responsibility for character may be that it makes the notion of identification redundant. In other words it may seem that being a morally responsible self is both a necessary and sufficient condition for moral responsibility for character. It may seem that as morally responsible selves we are automatically morally responsible for our characters. For surely if we are the principal authors of our self-narratives, then because our characters are part of our self-narratives we are morally responsible for our characters? Surely being the principal author of something makes you morally responsible for it? The answer is 'no, not necessarily'. The reason lies in the difference between knowing that something is part of us and really wanting something to be part of us. As morally responsible selves, we know our selves to a large extent. Because we are their principal authors, our self-narratives are to a large

extent formed by our self-representations and self-interpretations. But this does not necessarily mean that we are morally responsible for our characters. Even if we see our characters as being part of or identical with our self-narratives, knowing them does not make us morally responsible for them. We have to acknowledge that many things about us are part of who we are, without our really wanting these things to be part of us. In accordance with what I have said in chapter two, in order to be morally responsible for our characters, or for particular traits, we must identify with them. We must really want them to be our characters or traits.

A self-representation is essentially of the form 'X is part of who I am'. Individuals who are the principal authors of their self-narratives have self-narratives that are predominantly constituted by these self-representations. Individuals who are not their principal authors have self-narratives that are predominantly constituted by other people's representations of them. Such representations are essentially of the form 'X is part of who you are'. An identification in contrast to a representation is of the form 'X is part of who I really want to be'. So in order to be morally responsible for X, on my account, who I am must be predominantly constituted by self-representations and X must be part of who I really want to be. It is a matter of both being a morally responsible self and my identifying with my character.

4. Conclusion

I conclude that virtue ethics can include a notion of moral responsibility for character. The first condition for being attributed moral responsibility for our characters is that we are, to a significant extent, morally responsible selves. Being a

morally responsible self is a matter of being the principal author of a self-narrative. This is a matter of being the most persistent, or otherwise dominant, author amongst a diverse array of co-authors. The second condition is that we identify with our characters. The extent to which we identify with our characters determines the extent to which we are morally responsible for them. This is a matter of the extent to which our second-order volitions for them (or for the relevant primary desire for them) resonates throughout our, potentially endless, series of higher-order desires. Therefore we are morally responsible for our characters, or particular traits, when (1) we are to a significant extent morally responsible selves and (2) we identify with our characters, or traits.

References

- Beauchamp, T. (1995). 'What's so special about the virtue?' In E. Shelp (Ed.). *Virtue and Medicine: Explorations in the Character of Medicine*. Dordrecht: Reidel.
- Blum, L. (1994). *Moral Perception and Particularity*. New York: Cambridge University Press.
- Christman, J. (Ed.). (1989). *The Inner Citadel: Essays on Individual Autonomy*. New York: Oxford University Press.
- Dahlbom, B. (Ed.). (1993). *Dennett and His Critics: Demystifying the Mind*. Oxford: Blackwell.
- Dennett, D. (1978). 'Mechanism and responsibility'. In T. Honderich (Ed.). *Essays on Freedom of Action*. London: Routledge and Kegan Paul.
- Dennett, D. (1984a). *Elbow Room: The Varieties of Free Will Worth Wanting*. Oxford: Clarendon Press.
- Dennett, D. (1984b). 'I could have done otherwise – so what?' *Journal of Philosophy* 81: 553-565.
- Dennett, D. (1991a). *Consciousness Explained*. Boston: Little, Brown and Company.
- Dennett, D. (1991b). 'Real patterns'. *Journal of Philosophy* 88: 27-51.
- Dennett, D. (1993). 'Back from the drawing board.' In Dahlbom 1993.
- Double, R. (1991). *The Non-reality of Free Will*. New York: Oxford University Press.
- Dworkin, G. (1989). 'The concept of autonomy'. In Christman 1989.
- Fischer, J. (1988). 'Responsiveness and moral responsibility.' In Schoeman 1988.
- Fischer, J. and Ravizza, M. (Eds.). (1993). *Perspectives on Moral Responsibility*. New York: Cornell University Press.

- Flanagan, O. and Rorty, A. (Eds.). (1990). *Identity, Character and Morality*.
Massachusetts: MIT Press.
- Frankfurt, H. (1989a). *The Importance of What We Care About*. Cambridge:
Cambridge University Press.
- Frankfurt, H. (1989b). 'Alternate Possibilities and moral responsibility.' In Frankfurt
1989a, 1-10.
- Frankfurt, H. (1989c). 'Freedom of the will and the concept of a person.' In Frankfurt
1989a, 11-25.
- Frankfurt, H. (1989d). 'Three concepts of free action.' In Frankfurt 1989a, 47-57.
- Frankfurt, H. (1989e). 'Identification and externality.' In Frankfurt 1989a, 58-68.
- Frankfurt, H. (1989f). 'The importance of what we care about.' In Frankfurt 1989a,
80-94.
- Frankfurt, H. (1989g). 'What we are morally responsible for.' In Frankfurt 1989a, 95-
103.
- Frankfurt, H. (1989h). 'Identification and wholeheartedness.' In Frankfurt 1989a, 159-
76.
- Frankfurt, H. (1989i). 'Rationality and the unthinkable.' In Frankfurt 1989a, 177-190.
- Greenspan, P. (1988). 'Unfreedom and responsibility.' In Schoeman 1988.
- Hartz, G. (1990). 'Desire and emotion in the virtue tradition.' *Philosophia*. 20:145-65.
- Kaufman, A. (1972). 'Responsibility, moral and political.' In *The Encyclopedia of
Philosophy*. P. Edwards (Ed.). New York: Macmillan and The Free Press.
- Klein, M. (1995). 'Responsibility.' In *The Oxford Companion to Philosophy*. Oxford:
Oxford University Press.
- Kolm, S. (1981). 'The Buddhist theory of "no-self."' In J. Elster (Ed.) *The Multiple*

- Self*. Cambridge: Cambridge University Press.
- Kruschwitz, R. and Roberts, R. (Eds.). (1987). *The Virtues: Contemporary Essays on Moral Character*. California: Wadsworth.
- Kupperman, J. (1991) *Character*. New York: Oxford University Press.
- Marais, E. (1989) *The Soul of the Ape. The Soul of the White Ant*. Johannesburg: Jonathan Ball Publishers.
- MacIntyre, A. (1981). *After Virtue: A Study in Moral Theory*. London: Duckworth.
- Nagel, T. (1993). 'Moral luck'. In D. Statman (Ed.). *Moral luck*. Albany: State University of New York Press.
- Nussbaum, M. (1986). *The Fragility of Human Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. Cambridge: Cambridge University Press.
- Nussbaum, M. (1992). 'Virtue revived'. *Times Literary Supplement* 3.7.
- Oakley, J. (1996). 'Varieties of virtue ethics'. *Ratio* 9:128-52.
- Reber, A. (1995). *The Penguin Dictionary of Psychology*. London: Penguin.
- Rorty, R. (1993). 'Holism, intrinsicity and the ambition of transcendence'. In Dahlbom 1993.
- Ross, D. (unpublished M. S.). 'Rainforest realism'.
- Schoeman, F. (Ed.). (1988). *Responsibility, Character and the Emotions: New Essays in Moral Psychology*. Cambridge: Cambridge University Press.
- Slote, M. (1980). 'Understanding free will.' *Journal of Philosophy*. 77:156-51.
- Stocker, M. (1976). 'The schizophrenia of modern ethical theories'. *Journal of Philosophy* 73:453-466.
- Stump, E. (1993). 'Sanctification, hardening of the heart, and Frankfurt's concept of free will.' In Fischer 1993.

- Taylor, C. (1976). 'Responsibility for self'. In *The Identities of Persons*. A. Rorty (Ed.). Berkeley: University of California Press.
- Taylor, Charles. (1994). *Sources of the Self: The Making of the Modern Identity*. Cambridge, Massachusetts: Harvard University Press.
- Thalberg, I. (1989). 'Hierarchical analyses of unfree action'. In Christman 1989.
- Watson, G. (1989). 'Free agency'. In Christman 1989.
- Watson, G. (1990). 'On the primacy of character'. In Flanagan 1990.
- Watson, G. (1993). 'Responsibility and the limits of evil: variations on a Strawsonian theme.' Fischer 1993.
- Williams, B. (1983). *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press.
- Williams, B. (1985). *Ethics and the Limits of Philosophy*. Cambridge, Massachusetts: Harvard University Press.
- Wolf, S. (1988). 'Sanity and the metaphysics of responsibility'. In Schoeman 1988.
- Wolf, S. (1990). *Freedom within Reason*. New York: Oxford University Press.

