AN ESSAY ON PERSONAL IDENTITY AND RESPONSIBILITY

by

BANA BASHOUR

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy,

The City University of New York

2008

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Catherine Wilson

Date	Chair of Examining Committee

John Greenwood

Date	Executive Officer

John Greenwood

Michael Levin

Catherine Wilson
Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

AN ESSAY ON PERSONAL IDENTITY AND RESPONSIBILITY

by

Bana Bashour

Adviser: Professor John Greenwood

In the personal identity literature it is held that a later person remains responsible for an earlier action if and only if that person is identical to the one who was originally responsible for that action. In this dissertation, this is the claim that I will challenge. I will argue that a person may be responsible for an action even if they are different from the person who performed it. In other words, the thesis I will try to defend is that conditions of personal identity over time do not coincide with conditions of responsibility over time. I will argue that given that a being is a person who committed an act and was responsible at the time, that being will remain responsible for that act as long as that being remains **a** person. I will start off by presenting the different conditions of personhood, and then propose the outlines of a view of moral responsibility. I will discuss conditions of personal identity over time as well as conditions of moral responsibility over time. I hope to show that they are not equivalent, and bring up cases where a later person is responsible for an act

of a different (earlier) person. I will turn to cases from psychology of Multiple Personality Disorder and discuss synchronic multiple personhood and moral responsibility in these cases. I will look at especially puzzling examples that might cause problems for my account and deal with other possible problems and objections to this account.

Acknowledgements

I would like to begin by thanking my parents, Najla Nuseir and Maan Bashour. My mother, Najla, started writing her own dissertation during the Lebanese civil war, while teaching at the American University of Beirut, running her own company which makes educational toys, being involved in more charities than anyone can count, and taking care of five children. She always made it seem like everything was possible. My father, Maan, is still the same person that he was when he was young, he has never let anything get to him and change the ideals he lived by. He taught me that you should always stand up for what you believe in.

I would also like to thank John Greenwood for being everything any student could possibly want in an advisor. He was encouraging and supportive but most importantly, he was critical and challenging. He made writing the dissertation fun and inspiring. Not many advisors sit down with their students and look over the dissertation line by line.

Another committee member that I am deeply indebted to is Michael Levin. His comments were invaluable, and he was supportive throughout my graduate student years. Despite our philosophical disagreements, I cannot thank him enough for all of the conversations we've had.

I would also like to thank other professors from the Graduate Center, Nickolas Pappas, Richard Sorabji and Catherine Wilson for their encouragement and support. They gave me many extremely helpful comments that I will certainly use in my future work.

Had it not been for my siblings and their spouses who spoiled me during my vacations, I may have gone insane, so I would like to thank Rabi, Lama, Ahmad and Khaled Bashour. I would also like to thank Ameen Hannoun and Bushra Abdel Samad.

Next, I would like to thank members of the philosophy department at the American University of Beirut who were supportive during the semester I spent there writing this dissertation. Most importantly, I would like to thank Muhammad Ali Khalidi for encouraging me to go to graduate school in the first place and who inspired me to write this dissertation during a conversation at his kitchen table in one of my visits to Beirut. I would also like to thank Saleh Agha for getting me interested in philosophy in general, and the issue of personal identity in particular, in his introductory courses.

My friends from the Graduate Center, Ellen Fridland and Jean-David Lafrance, not only offered me support but also indulged me with invaluable conversations on my dissertation. I would like to thank them whole-heartedly.

I would also like to thank the following students for their friendship and support: Michael Benitez, Damien Dupont, Kevin Murtagh, Andrew Porter, Salas Sanchez-Benassar and James Snyder. I would also like to thank all the students who were present at the Graduate Student Colloquium; because of their questions I was able to shape my account more clearly. Also, I would like to thank the countless other graduate students and Faculty members at the Graduate Center with whom I engaged in numerous interesting and challenging philosophical conversations.

And I could not have done this without my other friends, my friends in New York: Sireen Hajj, Maher Khani, Reina Arakji, Nada Hamadeh, Fadi Kabalan, and Fadi Bardawil; and my friends overseas: Ghenwa Hayek, Zeina Halabi, Sherine Shallah and Samia Naccache.

Last but not at all in the least, I would like to thank Bashshar Haydar, for being everything that he is and for giving me a reason to finish.

Table of Contents:

Chapter 1 Introducing the Issue

It's 1977 and Zein is a heroin addict. She will lie, steal and even kill just to make sure she gets her fix. Some situations come up in which she does lie and steal. One day she even kills a man who refuses to give her his money when she threatens him. At the time she is perfectly aware of what she is doing, and given that she enjoys her lifestyle, she does not let anyone get in her way.

Thirty years later, it's 2007, and Zein has been clean for over twenty five years. She is a loving mother to two children who are in their last years of high school, and she is a reliable member of the community. She volunteers for organizations that rehabilitate drug addicts and get them off the streets. She gives occasional lectures about her days as an addict and hopes to inspire young people to quit or never start taking drugs. She is a soccer mom who is also a prominent member of the community.

However, the police find new evidence for an unsolved homicide in 1977, and are able to link the charitable soccer mom with the murderous drug addict. Now people in the community have to answer the following question: Is Zein, the loving wife and mother, still responsible for the heinous crime that was committed in 1977? Is she still the same person as that drug addict?

1. The Question

When Locke discussed the notion of personal identity in chapter XXVII of book II of his "Essay Concerning Human Understanding"[1], he introduced

[1] Locke, John *Essay Concerning Human Understanding* Ed P.N. Nidditch (Oxford, Oxford University Press, 1979)

the idea of a person in the forensic sense of the term. He was interested in the issue of moral responsibility, and according to him, a person is the type of being we can hold responsible for its actions. So we can say that responsibility is a specific type of relation between a specific type of being (a person) and a specific type of event (an action). He then gave a theory of what makes up a person and described how that type of being persists over time.

In the personal identity literature it is held that a later person remains responsible for an earlier action if and only if that person is identical to the one who was originally responsible for that action. In this dissertation, this is the claim that I will challenge. I will argue that a person may be responsible for an action (or state of affairs) even if they are different from the person who performed it (and who was responsible for it at the time). So although Zein, the soccer mom, is a different person from Zein, the drug addict, the soccer mom is still responsible for the murder that the drug addict had committed thirty years earlier. In other words, the thesis I will try to defend is that conditions of personal identity over time do not coincide with conditions of responsibility over time. I will argue that given that a being is a person who committed an act and was responsible at the time, that being will remain responsible for that act as long as that being remains **a** person.

In order to put forth my argument, I will start off by presenting the different conditions of personhood, and then propose the outlines of a view of moral responsibility. I will discuss conditions of personal identity over time as

well as conditions of moral responsibility over time. I hope to show that they are not equivalent, and bring up cases where a later person is responsible for an act of a different (earlier) person. I will turn to cases from psychology of Multiple Personality Disorder and discuss synchronic multiple personhood and moral responsibility in these cases. I will look at especially puzzling examples that might cause problems for my account and deal with other possible problems and objections to this account.

2. Historical Background

Although Locke is thought to be the first philosopher who discussed personal identity over time in relation to responsibility over time, it was Leibniz before him who was the first to connect the two. Since he was a Christian, Leibniz wanted to salvage the notion of responsibility in the after-life. I should not be rewarded or punished for something if I am not the same as the person who did it. According to Leibniz what is necessary for being the same person is my ability to recall what I did. Following from this, it is grossly unjust to reward or punish someone who doesn't have any memory of their past. Further, lack of memory of an earlier person's past makes the later person different from the earlier one. In the 'Discourse on Metaphysics', Leibniz says: *"But the intelligent soul, which knows what it is, and can say that word 'I', which says so much, not only metaphysically remains and subsists more than the others, but also morally remains the same and constitutes the same*

person. For it is memory, or the knowledge of this 'I', which renders it capable of punishment and reward."[2]

He goes on to give the famous king of China thought experiment, where he asks you to imagine that someone has given you the opportunity to become the king of China but only on the condition that all your memories of your past are erased. He concludes that no one would accept that because in some sense this "I" ceases to be the same "I" once all the memories are erased. You would not particularly care about that new person because it is no longer you. After your memories are erased you no longer exist, and the person is incarnated in your body, even if he were to be crowned the king of China would not be anyone that you have special concern for. It would certainly not be the same concern that you have for yourself (without any erasing of memories) in the future. This is supposed to show that we need the sameness of consciousness, or conscious memory of the past, for identity of this "I" we use.

One can see the connection that Leibniz makes between memory and responsibility. He believes that it is counterintuitive for us to hold that we ought to punish or reward anyone who did not have any memory of past actions that they are being punished or rewarded for. Leibniz also ties responsibility to the self, and suggests that you are not the same person as the one who had performed an action if you have no recollection of that

---

[2] Leibniz, Gottfried "Discourse on Metaphysics" Tr, R.S. Woolhouse and Richard Francks, *G.W. Leibniz Philosophical Texts* (New York, Oxford University Press, 1998) p. 86

earlier person's past. It seems clear here that he ties both the responsibility and the self-conscious memory to the continuing self.

In the 'Essay Concerning Human Understanding', Locke discusses the conditions of identity over time for many different beings. He says that for animals and plants, a later being is identical to an earlier one if and only if it "partakes in the same life"[3]. This means that my cat Puffy today is identical to the cat I had when I was in high school if and only if this later Puffy is physically continuous with the earlier Puffy and has remained a cat throughout (had the same functional organization and internal processes, etc…). The same applies to other sorts of animals and plant. Since humans are animals too, then the same conditions that make an animal identical to itself over time apply to humans as well. A human being is identical to an earlier one if and only if this later human being is physically continuous with the earlier human being, and has remained a human being throughout. However, Locke distinguishes human identity from personal identity. Although humans are animals, they are a special kind of animal since they are ones which we make moral judgments about. He introduces the notion of a person in the forensic sense of the term. A person is a type of being that we can hold responsible for its actions. Locke argues that the conditions of personal identity over time are different from the conditions of human identity over time.

---

[3] Ibid p. 443

To illustrate this difference, one can think of the following analogy:[4] Consider a bronze statue of Napoleon that is placed in the middle of a courtyard. The courtyard owner's daughter, Leena, decides that she would like to have a statue of Britney Spears instead of Napoleon. Of course her father does not share her desire. So Leena gradually changes this statue little by little until the same bronze makes a statue of Britney Spears. It is easy for us to say that this later statue is different from the earlier one, while maintaining that the later bronze is the same as the earlier bronze. In this case, there is nothing over and above the bronze that makes up the statue, but the fact that two statues are made of the same bronze does not necessarily mean that they are the same statue. This means that the same physical object, if picked out in different ways, might satisfy different conditions of identity over time. In the same way, an animal may still be made of the same material but cease to be the same animal, or even an animal at all. For instance, if that animal dies, it is still made of the same material but is no longer the same animal. We can pick out the animal by talking about the cells that make it up, or by talking about something more related to its functional organization and internal processes. What really makes a later thing identical to an earlier thing depends on what you are picking out in that thing. So the identity of a human being is different from the identity of the cells that make up that human being, although there is nothing over and above these cells that makes up the human being. Now to get back to our discussion of persons and humans, we

[4] This example is similar to the one used by Peter Geach when discussing relative identity in 'Identity' Review of Metaphysics, vol. 21, pp. 3-12, September 1967

can see why we don't need to posit immaterial souls or new entities to be able to make this distinction between a person and a human being. It could be that the person is only made of a body, but what counts for this person's being the same person over time is different from what makes this body the same body over time. What Locke was more concerned with, and what I am interested in for the purposes of this dissertation is what makes a person the same over time, and not what makes a human being the same over time. So our discussion is about what makes one the same being that can be held responsible for its actions over time.

Locke's view of what makes someone the same person over time was similar to Leibniz's but more elaborated. He believed that a being is a person at any given moment if that being is conscious of itself as itself. And what makes a later person identical to an earlier person is that they have the same consciousness or the same self conscious memories of the earlier person's actions:

"For since consciousness always accompanies thinking, and it is that that makes everyone to be what he calls self, and thereby distinguishes himself from all other thinking things: in this alone consists personal identity, i.e. sameness of a rational being. And as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then, and it is the same self with this present one that now reflects on it, that that action was done."[5]

---

[5] Locke, John *Essay Concerning Human Understanding* Ed P.N. Nidditch (Oxford, Oxford University

I will discuss Locke's account of what makes a person the same over time in more detail in the next chapter. I will also discuss some of the main objections in later sections as well. Two noteworthy objections are those of Joseph Butler and Thomas Reid. The one presented by Butler is that Locke's account is circular, and that his criterion of sameness of consciousness already assumes personal identity over time.[6] I will discuss this objection in the next chapter. Reid's objection is that his account fails to satisfy the transitivity of identity.[7] Reid claims that according to Locke's account, we can have three persons, the first of which is identical to the second, the second is identical to the third, but the first not identical to the third. This is obviously problematic for Locke's account because we think of identity as transitive. I will leave a detailed discussion of this objection for chapter 4.

One thing it is important to point out concerns Locke's main contribution. Henry Allison noted that the reason Locke's account is important is not that he gave a good or nearly exhaustive account of the necessary and sufficient conditions for remaining the same person over time, but that he separated the metaphysical question from the forensic/moral one.[8] The former of these questions, according to Locke, has to do with bodies and

---

Press, 1979) p. 335

[6] Butler, J. 'Of Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 99-106

[7] Reid, T. 'Of Mr. Locke's Account of Our Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 113 –118

[8] Allison, H. 'Locke's Theory of Personal Identity: A Re-examination' *Journal of the History of Ideas*, vol. 2, pt. 7, pp. 41-58, January-March 1966

substances persisting over time. However, the latter has to do with consciousness and awareness of oneself, which is what he considers to be essential for personal identity in the moral sense. Although in providing an account for personal identity he gives a metaphysical story, his primary concern is a normative one. This is what his distinction between human identity and personal identity over time tells us. Giving necessary and sufficient conditions of human identity over time does not give us answers to questions regarding our moral concerns. Discussions of personal identity are helpful in that respect. We cannot talk about responsibility (which is my concern in this work) unless we are talking about personal identity. Locke recognized that and made a very helpful distinction that enabled us to discuss this problem without denying that humans are animals, and that conditions of persistence over time for animals in general apply to humans. This is a distinction that I will adopt and will rely on for my account.

Anthony Flew argues that although Locke saw the importance of the problem of personal identity for responsibility, we might not need to completely resolve this issue in order to ground responsibility.[9] Flew argues that the big source of trouble for Locke is that he introduced the idea of 'person' and distinguished it from all else, leaving him with nothing but an account that has to be a circular one. I will argue in the next chapter that it is only because of the circular assumptions concerning self-consciousness that

---

[9] Flew, A. 'Locke and the Problem of Personal Identity' *The Journal of the Royal Institute of Philosophy*, vol. 2, pt. 6, pp. 53-68, January 1951

this problem arises. It is not the distinction between human identity and personal identity that makes Locke's account fail, but it is the specific conditions that he gives for personal identity over time. However, there I do agree with Flew's idea that it might be possible to distinguish between remaining the same person over time and remaining responsible for a certain action over time. So by answering questions of whether a later person is identical to an earlier person we do not answer the question of whether this later person remains responsible for the earlier person's actions. I am sympathetic with that criticism. Even though persons are the only type of beings that we can hold responsible for their actions, this does not guarantee that remaining the same person over time is a necessary and sufficient condition for remaining responsible for an act over time. I will discuss this in more detail in Chapter 4.

3. Some contemporary theories

Some philosophers agree with Locke and claim that it is only psychological continuity that is needed for personal identity over time, while others rely on bodily continuity. Among those who defended theories based on psychological continuity, some defended different versions of Locke's memory criterion while others defended other types of psychological continuity. Some claimed that neither bodily nor psychological continuity on its own is enough, but what is needed is a mixture of both. In this section, I will briefly discuss some versions of these different theories and look at what

each of them would say about remaining responsible for a given act over time.

a. Psychological continuity

Memory

Although the two objections mentioned above (and that will be discussed later) were considered fatal for Locke's account, some philosophers attempted to save the memory criterion from these objections. The best version of a memory based theory is presented by H.P. Grice.[10] Grice gave up talk of having the same consciousness and replaced it with talk of memory. He claimed that this substitution avoids the charge of circularity because there is no reference to self-consciousness (Butler's problem with Locke's account). By discussing memory he thought he had avoided this objection. Next, he introduced the term "total temporary state" which refers to "*all the experiences any one person is having at any given time*"[11]. For example, if I am eating a candy bar at a moment, while thinking about my dissertation and staring at my computer, there occurs a total temporary state containing as elements the thought of my dissertation, the taste of the candy bar and the sensations from my computer. In other words, the total temporary state includes as elements all the mental states that I am in at that moment. Memories are also members of these total temporary states. So for instance, in the example above, I can have the memory of eating oranges at the same

---

[10] Grice, H.P. 'Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 73-98

[11] Ibid p. 86

time as all these. This past memory is part of that total temporary state. He went on to define personal identity over time in terms of these total temporary states:

"…in a series of total temporary states belonging to one person, every t.t.s which is a member of that series will contain as an element a memory of some experience which is an element in the temporally preceding member of the series; in a series of total temporary states not belonging to one person this will not be the case."[12]

So on this new account, it is not sharing a particular memory that makes one identical to another, it is sharing the series of total temporary states. Even if the later total temporary state shares nothing with the earlier one, but they both share some states with a total temporary state in the middle, then they all belong to the same person. A total temporary state only has to share something with one other total temporary state for it to be part of the same series, and hence belong to the same person. This helps him deal with Reid's objection, because on this account transitivity holds. By talking about total temporary states, in which each total temporary state contains elements of past temporary states it is difficult for the transitivity objection to go through. The total temporary state at t1 does not have to share anything with a total temporary state at t3 for them to belong to the same person. Consider a case in which we have a total temporary state at t1 that contains a memory that it shares with a different total temporary state at t2. Now at t3,

---

[12] Ibid p. 86

we have a total temporary state that shares nothing at all with t1, but it shares some states with t2. This is enough for them to belong to the same person

However, one objection that arises for Grice's account concerns false memories. Memory is a success term. If one (successfully) remembers an event, then he must be the same as the person who was present at the time the event occurred. Grice tried to answer this objection by giving a causal account of memory, and instead of having the account rely on psychological states on their own he gives a causal story for how these psychological states came about. However, he needs more than simple memory to make his account work. By attempting to answer what seems to be a fatal objection, Grice radically modified the memory account. Instead of talking about memories, he switched to talk of successful memories, which makes him fall back into the circularity objection.[13]

Although Grice himself did not discuss how this affects continued responsibility over time, we can assume that since his account is similar to Locke's then he also considers remaining the same person to be a necessary and sufficient condition for remaining responsible for an act over time. So according to Grice, necessary and sufficient conditions for being the same person over time are the same as the ones for remaining responsible for an act over time.

Parfit

---

[13] See Perry, J. 'The Problem of Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 3-32

The most influential account that is based on psychological continuity is that of Derek Parfit.[14] Parfit believes that people are interested in the question of personal identity because of concern for future selves. His conclusion is that what matters for survival is not identity, but a different relation, the R-relation. He argues that since identity is a one-to-one relation, but what matters to us can be a one-to-many or even a many-to one relation, then identity must not be what matters for survival. This is illustrated by examples of fission, where a person splits into two persons. In these cases, we cannot say that these two resulting persons are identical to the original person, because that would make them identical to each other, which is impossible. He claims that in these cases, even though we don't have identity, we still have what matters for survival. If someone told you that starting tomorrow there will be two persons that have your same psychological features and memories of your past, this certainly seems to be much better for you than if he were to tell you that you will die tomorrow. So most of us prefer fission to death, and that is because fission still gives us what matters for survival, although it does not give us identity. So what matters in survival is different from personal identity. Parfit also argues that cases of fusion teach us that what counts is a matter of degrees. He concludes that what matters for survival consists in psychological continuity. To make sense of talking about psychological continuity in a non-question begging way (and to avoid one of the major objections to Locke), he

---

[14] Parfit, D. 'Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 199-226

introduces what he calls “q-memory” (or quasi-memory). I q-remember an experience A if and only if the following three conditions apply. First, I have a belief about this past experience A that seems to be like a memory belief. Second, someone did have such an experience. And third my belief is dependent on this experience in the same way memory beliefs are dependent on past experiences. He claims that we can apply these criteria the same way to q-intentions, q-recognizing, q-promising, and most importantly for our purposes, q-responsibility. These new terms allow us to talk about psychological states in a non-circular and “impersonal” way. It also enables us to talk about what really matters without talking about identity. So what matters for survival is not identity, but psychological continuity.

Since my view of personal identity over time will also focus on psychological states, I will agree with Parfit on some issues. I will discuss these issues in more detail in Chapter 4. However, he does tie responsibility to what matters for personal identity more than I would like to. He argues that what matters (or this R-relation) also grounds responsibility. He presents an example of a man over ninety who rightfully won the Nobel Peace Prize. This man confesses that he beat up a police officer in a drunken fight when he was twenty. Parfit believes that this man is no longer responsible for that action, because he is so different from the one who actually did it.[15] So even Parfit makes the connection stronger than I would like to do. I will argue that the Nobel Peace Prize winner is still responsible for what the twenty year old man

---

[15] Parfit, D. *Reasons and Persons*. (New York, Oxford University Press, 1984)

had done, even if he is a different person. This is one of my major disagreements with Parfit.

b. Bodily continuity

Some philosophers claim that, just like human identity, personal identity requires physical continuity. These philosophers argue that it is a mistake to discuss psychological criteria on their own without discussing physical criteria. They conclude that it is physical continuity rather than the psychological continuity that is important for continuing persons. Bernard Williams argues for this position. Defenders of psychological continuity views usually appeal to thought experiments in which you are asked to imagine cases in which there is psychological continuity but no physical continuity. Most of these suggest that what is important is psychological continuity rather than physical continuity. Williams argues that these thought experiments are always phrased in question-begging ways.[16] If the thought experiments were phrased differently, then they might yield a reverse result. He gives the following example: Suppose you tell someone that he will be tortured tomorrow. You then try to comfort that person by telling them that his memories and his character will be completely erased. This prospect doesn't seem to alleviate his fear or concern of what will happen to him. In fact, if anything, this might make the person even more worried. We might be inclined at this point to say that physical continuity is necessary for survival. So if the thought experiment is phrased this way, then we will have completely different intuitions about the

---

[16] Williams, B. 'The Self and the Future' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 179-198

necessity of psychological continuity. The idea behind this example is to show that the thought experiments that defenders of psychological criteria invoke are misleading, and it is the description of the thought experiment that pushes people to adopt views other than those of bodily continuity. If phrased differently, a though experiment might make one find bodily continuity theories preferable. For that reason, he suggests that personal identity over time consists in physical continuity rather than psychological continuity. Since the thought experiments that his opponents use are of no value, physical continuity seems like the better of the two options. Personal identity is nothing over and above human identity.

Williams and other proponents of the physical view argue that you are the same person as the one who committed an act if and only if you are physically continuous with that person. This means that you are responsible for the actions of the person that you are physically continuous with. So the conditions of personal identity over time are the same as the conditions of remaining responsible for an act over time. This is, again, the claim that I will disagree with.

c. Hybrid theories

Philosophers who argue for the psychological or the bodily continuity views believe that it is either exclusively physical or exclusively psychological features of persons that are essential for personal identity over time. Since most contemporary philosophers are not Cartesian dualists, some have argued that the separation between physical and psychological conditions

might not be as clear as it sounds. In fact, if you are a materialist about the mind, you will have to identify the psychological features with some physical ones. John Perry is one of the philosophers who argue along this line.[17] Perry agrees with Locke and other theorists that memory and psychological continuity are essential for personal identity, but he points our attention to the force of some recent arguments and studies which show that it is important to have physical continuity to have identity over time. He goes on to argue that these claims do not conflict, but together constitute the right theory of personal identity. He proposes the view that personal identity consists in having the same brain, which is the physical entity that holds all the psychological features that are interesting for us.

Perry argues that radical changes in personality cannot make a being change from one person into another because it is a contingent fact that persons change their personality without becoming different persons. He goes on to say that I will still be concerned with my future self even if I knew I was going to have a drastic change of personality. He claims that this destroys the view that personal identity over time is based upon particularities of your psychology. In chapter 4, I will discuss this objection in detail. I will argue that it begs the question against my account, and that although you become a different "person" over time, you still remain responsible for your earlier actions,

4. Distinguishing selves from persons

---

[17] Perry, J. 'The Importance of Being Identical' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 67-90

Locke's theory of personal identity over time relies on the idea of self-consciousness. As mentioned above, what makes someone the same person over time is that he continues to have the same consciousness. Locke, as well as most authors after him, identified persons with selves, and used the terms interchangeably. Selves are considered to be subjects of experience. Although it may be true that only subjects of experience can be considered persons in our sense of the term and that it is only sameness of this subject that makes one the same person over time, this thesis should be argued for and not taken for granted. In this dissertation I am not concerned with the self or the subject of experience. I am concerned with a specific type of being, namely the person, and how this specific type of being persists over time. I will not make any assumptions about whether this person has to be the subject of experience or not, because I don't think it is necessary for my account. Although I will deal with self-consciousness and discuss it as one of my conditions, it will have a very specific meaning which has very little to do with subjects of experience. There is a dispute concerning whether selves are just persons engaging in reflexive reference, but I will not take a stand on this issue. I will only use the term 'person' in my discussion, and by 'person' I mean the specific term that Locke distinguished from the term 'human'.

The reason I would like to make this distinction is that there are many objections to the idea of a self, and many authors are skeptical of its existence or its persistence over time. I do not think that any of these objections apply to persisting persons over time. A general approach that

many of these objections follow is one presented by David Hume.[18] Hume denied the persistence of objects over time in general, but specifically the persistence of selves over time. Since we never have a direct experience of the self, then there we have no reason for believing in such a thing. We are never in direct contact with the self, and therefore there is no need to assume that there is one unified thing called a self that persists over time. We only have access to a collection of successive perceptions none of which uniquely identifies the self. I will not attempt to discuss whether this objection to the persistence of a self over time is valid or not. Even if this objection holds, it poses no threat to continuing persons over time. Personal identity over time has to do with a type of being persisting over time, regardless of whether or not there is a distinctive self and that self persists over time. What makes something a person is a normative question to which we cannot raise strict metaphysical objections. One does not have to assume that a person is identical to a self. It is only some accounts of personal identity that hold that view, and that is a claim that needs to be argued for rather than simply assumed.

One may object to this by saying that if selves were different from persons, can't we say that you don't need personal identity over time for one to remain responsible for an action, but you need identity of the self? In other words, an objection might arise concerning the fact that we still have something persisting over time and it is the identity of that thing (whatever

---

[18] Hume, D. *A Treatise of Human Nature* by Ed E. Mossner (London, Penguin Books, reprint 1969)

that thing is) that is required for remaining responsible.[19] I can respond to this by saying that if this self is not a person, then it is not the type of thing that can be held responsible for its actions. So it cannot be what persists for responsibility to remain. But if the self is identical to the person, then this objection does not arise. I must note that this objection can arise in a different and stronger version to my account in general, and I will deal with it in more detail in chapter 4.

I must here note that although selves and persons might be extensionally identical, they are intentionally different. We mean a different thing when we are talking about persons from when we are talking about selves. By selves we mean subjects of experience, whatever those are. By persons we mean the types of beings that can be held responsible for their actions, whatever those are.

So the subject of this dissertation is focused. It has to do with persons, what makes these persons the same over time, and what makes them remain responsible for an act over time. Since I am not concerned with the self, I am also not concerned with many of the questions concerning the self. I am not concerned with questions of self-interest, nor with responsibilities you have towards yourself. I will also not be discussing survival of the self, but I will be discussing survival of a person who is responsible for a given action. I should point out here that survival of a person does not imply survival of the same person. So in short, I will not be discussing what most philosophers discuss,

---

[19] I would like to thank Ellen Fridland for bringing this objection to my attention.

namely concerns for future selves and survival. I will only talk about persons, and whether they remain responsible for earlier actions that earlier persons had done.

5. Introducing my account

In this dissertation, I wish to argue that the conditions of personal identity over time are different from the conditions of continuing responsibility for an act over time. All of the philosophers mentioned above seem to take this equivalence for granted. I will challenge it. Although persons are the only types of beings that we can hold responsible for their actions, remaining the same person over time is not equivalent to remaining responsible for an action over time. There are different conditions that make a person responsible for an act over time, which I will discuss in Chapter 4.

A general outline of this account goes as follows. Personhood has to do with character, and the types of actions endorsed by that being. So if a person radically changes his character and a large number of things that he endorses, then he becomes a different person. Responsibility also has to do with types of actions endorsed by a being, but changing one's attitudes is not enough for them to stop being responsible for an action. A common case of one changing one's endorsements is when one regrets what one has done. Regretting what you did does not take away from the fact that you are responsible for that action, even if you are genuinely sorry. If you change a lot of the things you endorse, we may say that you have become a different person, but that is not enough for you to stop being responsible for an action.

So in the example above, Zein is a different person at 2007 from Zein in 1977, but Zein in 2007 is still responsible for the murder that happened in 1977. Although this claim sounds odd to the philosopher, I will argue that one may become a different person over time while remaining responsible for his actions. Although personal identity has to do with character, responsibility over time does not.

Chapter 2 Conditions of Personhood

It is natural to start our inquiry with a discussion of what it means for someone to be a person. Although the issue of identity over time is one that is essential for our purposes, we will leave the discussion of that for a later section. In this chapter I will discuss different proposed conditions for being a person at any one moment, and hopefully give an account of that. This will also help us with our account of personal identity over time, the subject of chapter 4.

1. Basic Conditions

Before discussing some conditions of personhood that are specific to my account, it is helpful to start with basic uncontroversial conditions. These conditions will hopefully be common to most views of what makes something a person. It is only the later conditions that are more controversial.

a. Embodiment

The first and least controversial condition for personhood is embodiment. Most philosophers today are not Cartesian dualists about the mind and do not believe that mental states exist outside the body. Therefore, very few will disagree that persons cannot exist without bodies. Also, since we are discussing beings that can be held responsible for their actions, they must be the types of beings that are able to perform actions. They must have some motor abilities. They must also be able to have some perception, and it is in virtue of these perceptions and produced (or potential) movements that

they are persons. So this condition specifies that no persons can exist without a body.[20]

Some philosophers have argued that this condition is what justifies the unity of agency which shows that there is only one person for every body. Aristotle defends this, arguing that although the soul might have different parts and different functions that these parts perform, there is only one body that can be moved.[21] So since there is only one body that can be set in motion, there is one agent that can act. Two conflicting desires cannot both get their way. So even if a person is made up of a collection of mental states, he still has to be a unified person. This may or may not have some implication for a person's persistence over time. However, for the purposes of this chapter, I would maintain the weaker thesis that in order for one to be a person at any given time, they have to be embodied.

One may ask the following question: Does it have to be one person per body? Can there be one body with more than one person? And can we have a person with more than one body? The answer to the first question is no. In chapter 5 I will discuss cases of multiple personality disorder in which I will argue that there is more than one person instantiated in the same body, which gives an affirmative answer to the second question. As for the third question, since the conditions that I will discuss later on in this chapter concern some intentional states, it would be difficult, if not impossible,

---

[20] Sidney Shoemaker argues in support of this condition in 'Embodiment and Behavior' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976)

[21] Aristotle, *De Anima* Book III Chapters 9 and 10

for two bodies to have numerically identical intentional states. Since it is these states that make one a particular person, then it is impossible to have one person with two bodies. But that will be clearer once we have discussed which particularities of one's psychology make one a specific person.

b. Sentience

Another basic condition for personhood is sentience. Candidates for personhood must be conscious on some basic level, which means that they must be able to respond to their environment. To use contemporary philosophical terminology, candidates for personhood must be capable of having creature consciousness, which is defined by David Rosenthal as the following:

"Being conscious in this sense is, roughly, the opposite of being asleep or knocked out; we describe a person or other animal as being conscious if it is awake and if at least some of its sensory systems are receptive in the way normal for a waking state. Otherwise we say it's unconscious."[22]

In a later section of this chapter, I will discuss a different kind of consciousness that is more controversial. However, the ability to be aware of one's surrounding is something that most creatures have, and persons are required to have that as well. In order for persons to act, they need to have some minimal awareness of their environment.

2. Dennett

---

[22] Rosenthal, David 'State Consciousness and Transitive Consciousness' *Consciousness and Cognition*, 2, 4 (December 1993), pp. 355-363

For the other conditions I will rely on Daniel Dennett's account[23] of personhood. I must note that Dennett and I are dealing with slightly different questions. He makes a distinction similar to the one I make between metaphysical personhood (selves) and moral personhood. He believes that his conditions are meant to be conditions for metaphysical personhood and not moral personhood. He adds that metaphysical personhood is necessary but not sufficient for moral personhood. I do not agree that metaphysical personhood is necessary for moral personhood but I do accept these conditions as necessary for personhood. This is why I rely partly on his conditions but I also add some other conditions in order to ground moral personhood. He discusses six different conditions of personhood. I will only deal with five of them in this section. I will leave another for the next section. I will group the first three of them (as he does) in the following section.

a. Intentional Systems

One basic condition of personhood is that the being in question must be an intentional system. We cannot discuss responsibility of an action performed by a non-intentional system, since it is impossible for a non-intentional system to act at all. Even if we were to attribute actions to these non-intentional systems, we don't really think it plausible to attribute any normative properties to a being that we cannot attribute intentionality to. In fact, much of the later and more controversial conditions will rely on particular intentional states. A lot of the ethical discussions have to do with the beliefs

---

[23] Dennett, Daniel 'Conditions of Personhood' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976)

and desires of the being in question. I will go over these discussions in detail in the next section, which will include an account of which of these intentional states are crucial for our purposes.

It might help here to look at Dennett's account of intentional systems.[24] Dennett gives three conditions for something to be an intentional system, all of which have to apply to persons. First, the systems in question must be rational. Second they must be beings to which intentional predicates are attributed. Third they must be ones towards which you can adopt the intentional stance.

In order for a being to be an intentional system, that being must possess rationality. Some philosophers, like Kant and Rawls, have argued that it is by virtue of their rationality that men are persons. Kant argued that being rational agents we recognize the fact that we must act in accordance with the categorical imperative. Kant grounds both personhood and his ethical theory in the rationality of humans.[25] John Rawls grounds personhood in rationality as well, and believes that it is this rationality that leads a person to recognize the principles of justice.[26] I will not be using as robust a notion of rationality as that of Kant or Rawls. By 'rationality' here I only mean basic deductive and inductive abilities. Rational beings are the types of beings that are capable of learning from past experience, and are able to perform

---

[24] Dennett, Daniel 'Intentional Systems' *Journal of Philosophy*, vol. 68, pp. 87-106, February 1971

[25] Kant, Immanuel *Groundwork for the Metaphysics of Morals* Tr. Thomas E. Hill Jr. and Arnulf Zweig (New York, Oxford University Press, 2002)

[26] Rawls, John , *A Theory of Justice* (Cambridge, Belknap Press, 1972)

somewhat complex tasks using this knowledge. Although one might want to attribute this weak sort of rationality to simple systems such as calculators and some computer programs, further conditions will rule out these systems as persons.

The second condition that Dennett sets is that candidates for personhood must have intentional or psychological predicates ascribed to them. One way to cash this out is to say that they are beings to whom you can attribute consciousness. I will not deal with consciousness in this section, but I will do so in a later one. All we need for now is to say that these beings are ones to whom we can attribute intentional states. This condition is related to the next one because in order for us to make such attributions we need to take a specific stance, namely the intentional stance.

A third condition for a system to be considered an intentional one is that it is one towards which we can adopt the intentional stance. This means that you explain this system's behavior by appealing to this system's intentional states. If adopting the intentional stance towards a system proves to be useful in predicting its behavior, then it is an intentional system. Cases you successfully adopt the intentional stance towards a system are ones in which:

'One predicts behavior in such a case by ascribing to the system the possession of certain information and by supposing it to be directed by certain

goals, and then by working out the most reasonable or appropriate action on the basis of these ascriptions and suppositions.[27]

There are of course other ways of explaining these systems' behaviors, like mechanistic and physical ways, but the intentional stance has proven to be very useful for certain beings. Adopting the intentional stance enables us to talk about the same phenomena in a different way, and that helps us shed light on the matters being discussed. We rely on the intentional stance every day. We attribute beliefs and intentions to other people and explain their behaviors accordingly. Some adopt the intentional stance towards animals, and consider animals to be intentional systems. You can even adopt it towards a plant, so you might consider the plant a low-grade intentional system.

One way to bring out this idea is to mention Kant's distinction between two different standpoints. Kant believes that there is a distinction between the way things really are and our knowledge of their appearance.[28] Some have called this the distinction between the theoretical and the practical standpoint.[29] You may describe one's actions in terms of neurons firing and biological happenings, but that does not help us understand reasons for the person's actions. We should be able to answer normative questions through

---

[27] Dennett, Daniel, 'Intentional Systems' p. 90

[28] Kant, Immanuel *Groundwork for the Metaphysics of Morals* Tr. Thomas E. Hill Jr. and Arnulf Zweig (New York, Oxford University Press, 2002)

[29] Korsgaard, C. 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit' from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p. 323-338

adopting a different standpoint (or stance) towards the agents we are discussing. We attribute certain intentions and desires to the agents who are acting, and we can make moral judgments based on those. This is a very similar idea to Dennett's intentional stance. For Dennett, the reason we adopt the intentional stance is pragmatic. Adopting such a stance might help us predict people's behaviors. For Kant, the reason we adopt the practical standpoint is in order to assess the moral worth of actions. Since I am interested in moral personhood, both Dennett's and Kant's accounts will be useful for me.

I will not attempt to give an account of what I believe mental states to be. I do not need to rely on any account of what these intentional states are, except to say that persons are the types of beings that you can attribute these intentional states to. Some may argue that Dennett's idea of an intentional system is too liberal because it doesn't require that the intentional systems actually have these intentional states, but only that it would be reasonable to attribute these states to them. He believes that considering something to be an intentional system is just pragmatic because it reliably helps us predict behavior. Many people disagree with this as a condition, and believe that it is not only necessary for the beings in questions to have intentional states attributed to them, but they should be required to have these intentional states as well. There must be something other than this predictive success to ground personhood. Since our concern is a normative one and not a metaphysical one, I will argue in the next chapter what we need to ground

responsibility is an ability to make these attributions. Although human beings do have intentional states, it is the attribution of these intentional states that explains holding people responsible. That is why we need to take the intentional stance, and not a physical or chemical one. I should note that this is only one condition for personhood, and certainly not the only one. The next few conditions should help clear up these doubts and address some of the concerns behind this objection.

It is important to note here that I do not need to be an attributivist about intentional states for my story to be a plausible one. I am not arguing that we do not have intentional states and that all we can say about these intentional states is that they can be attributed to us. Rather I wish to argue that both this ability to make such attribution and the fact that these attributions make sense when applied to us make us the kind of beings that hold others responsible and can be held responsible for their actions. The fact that we have intentional states is of course crucial, but we can explain responsibility better in terms of these attributions. Having intentional states will be a necessary condition for being a person as well, and I will discuss this while presenting the last condition.

b. Reciprocation

A fourth and related condition for beings to be persons is that they are capable of reciprocating the treatment mentioned in the third condition. In other words, they should also be able to treat others as persons.[30]

---

[30] In reference to the point mentioned above, for persons to be able to make such attributions, they

"This reciprocity has been rather uninformatively expressed by the slogan: to be a person is to treat others as persons."[31]

They should be able to attribute intentional states to others. Strawson gives a more informative account of this condition and argues that one cannot recognize one's own intentional states unless they have recognized these intentional states in others.[32] This entails that one cannot be a person unless he recognizes others as persons as well. This reciprocation is a necessary condition for personhood.

It is important to distinguish this condition of personhood from the prescriptive thesis proposed by Kant. Dennett mentions this as a requirement for something to be a person, while Kant argues that we ought to treat the humanity within a person as an end in itself and never as a means to an end.[33] I am not concerned with Kant's thesis, because this is not a proposal for an ethical theory. All we are concerned with are the conditions that make something the type of being that can be held responsible for its actions. I will argue later on in this chapter that is this condition in conjunction with the last one that ground persons' capacity to distinguish between right and wrong, which is what makes them moral agents.

c. Verbal Communication

---

certainly need to have intentional states.

[31] Dennett, Daniel 'Conditions of Personhood' p. 178

[32] Strawson, P.F. 'Persons' from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 104-116

[33] Kant, Immanuel *Groundwork for the Metaphysics of Morals* Tr. Thomas E. Hill Jr. and Arnulf Zweig (New York, Oxford University Press, 2002)

Another condition that Dennett discusses which I will deal with briefly is that the being must be capable of verbal communication. Some might argue that nothing can be an intentional system without having linguistic abilities. That is a controversial claim. I do not need to deal with this question for our current purposes, but I do need to address whether verbal communication is a necessary condition for personhood in our sense of the term.

Dennett argues that in order for one to be a moral agent, one has to be able to communicate with another, because that is what social contracts are based on.[34] I agree with him on this condition, but this needs a further argument. One cannot take the intentional stance, or attribute intentional states to others without having any linguistic abilities. To attribute a belief to someone else, you must be able to represent this belief. You must then recognize that belief in another. Both of these require a sophisticated conceptual framework that presupposes linguistic abilities. The last condition I will mention also requires that the candidates for personhood have linguistic abilities, because you need to have a complex conceptual framework in order to meet this condition.

d. Core Psychology and Distinctive Psychology

Of course, the conditions above do not by themselves give us a sufficient account of what it is to be a person, but they are definitely the starting point. It might be helpful here to draw on a distinction that Peter

---

[34] Dennett, Daniel 'Conditions of Personhood' p. 178

Unger[35] makes between a person's *core psychology* and his *distinctive psychology*. The core psychology consists in abilities and faculties that a person shares with other persons. Unger says:

"Among the mental capacities that it comprises, my dispositional psychology includes those that I share with all other normal humans, notably my capacity for conscious experience, my capacity to reason at least in a rudimentary way, and my capacity to form some simple intentions… We may call this group of capacities my core psychology…"[36]

The distinctive psychology consists in the person's particular combination of beliefs and desires that no one else can be presumed to have.

"Certain other aspects of my psychology I share with some normal humans but do not share with others, for example my (ostensible) memory of having tasted butter pecan ice cream. Still other aspects distinguish me from all other actual human beings, for example, my (ostensible) memory of having written the preceding chapter. Let us say that my distinctive psychology comprises both aspects of my psychology."[37]

He adds that the core psychology and distinctive psychology are both exclusive and exhaustive. The five conditions we have discussed so far comprise the core psychology of persons, but certainly not their distinctive psychology. Although Unger disagrees, the latter seems to be the one that is

---

[35] Unger, Peter 'The Physical View' from D. Kolak and R. Martin (ed.) *Self and Identity* p. 192-213 (New York, Macmillan Publishing Company, 1991)

[36] Ibid p. 193

[37] Ibid p. 194

more relevant for a person's identity over time. Unger claims that one's distinctive psychology changes too often for it to be what makes him the same over time. That is why maintaining the same core psychology is what is essential for personal identity over time. Since, as I will argue later on in this chapter, it is the person's distinctive psychology that explains his ability to distinguish between right and wrong, then that is what matters for personhood in our sense of the term. My concern is with responsibility and what makes someone the same responsible agent over time, and that is why the distinctive psychology might be more relevant. I will argue for that account of personal identity over time in chapter 4, but for now we need to define what makes someone a person at a given time. The core psychology makes one a person, while the distinctive psychology makes one a particular person. I will argue that what makes someone a particular person, which is essential for personhood, has to do with character, and that is determined by one's particularities, i.e. his distinctive psychology. I will discuss what this distinctive psychology consists in later on. It is this distinctive psychology that is more relevant to our normative concerns, and it has to do with specific types of intentional states.

e. Brain

In the previous chapter I mentioned that some philosophers believe that the necessary and sufficient conditions of personal identity include having the same brain. Before turning to a discussion of what one's distinctive psychology consists in, I must address the brain condition. Some, including

Unger, have argued that to be a person one must have a brain. Since only beings with brains can have a core psychology, only these beings can be persons. Unger also adds that in order for one to remain the same person over time, he has to keep the same brain and same core psychology. He notes that the brain condition is not a logically necessary condition. However, it is an empirical fact that the brain and nervous system (rather than any other organ) support and realize the intentional states. He explains:

"A person's physical parts and structures are important to her survival only insofar as they continue to support, and to realize, her basic psychological capacities"[38]

I do not disagree with Unger on this empirical fact, nor do I need to take a stand concerning what kind of things can have intentional states or a core psychology. All I need to do is point Unger's attention to the fact that I am dealing with a normative question, and that the distinction made earlier between taking the intentional stance and taking a different stance towards the phenomena takes care of this objection. It might be true that only beings with brains can have the core psychology required to be persons (functionalist objections to this aside), but that does not help us for our present purposes. If we discuss brains or other organs, we are no longer taking the intentional stance but are taking a physical or chemical stance. That is why this condition is of no interest to our current purposes. It may be that it is empirically necessary for a person to have a brain that realizes these psychological

---

[38] Ibid p. 195

capacities, but it is these psychological features that are crucial for our purposes. Whatever it takes to allow these features to exist is only relevant derivatively, because it realizes these features. We need to focus on what is important for us here which are these psychological features and not whatever realizes them. To use Kantian terminology, setting the brain as a condition for personhood would mean we are taking the theoretical standpoint rather than the practical one.

Now we need to turn to a discussion concerning which intentional states constitute one's distinctive psychology. This is the condition that Dennett mentions that I skipped in this section. I turn to it next.

3. Self Consciousness

a. Locke

Since we started our inquiry with Locke, it is natural to look at his account of personhood. I discussed Locke's view in the previous chapter. Let us now recall this view. Locke believed that a being is a person (in the forensic sense of the term) if that being is capable of being conscious of itself as itself.[39] In other words, this being has to have some sort of self awareness. It is the capacity for reflexive attribution or awareness that distinguishes moral agents from other creatures. If a being can make itself the object of its own thinking, then that being is a person. Since other species are incapable of such conscious rational thought, then they are not considered persons. Locke also concludes from this a view of personal identity over time. As I mentioned

---

[39] Locke, John *Essay Concerning Human Understanding* Ed P.N. Nidditch (Oxford, Oxford University Press, 1979)

in the earlier chapter, you are the same person as long as your consciousness extends. Another way of explaining this is to say that A at time t1 is the same person as B at time t2 if and only if B has self-conscious memories of A's experience (t2>t1). So what makes someone a person at a given time is that they are self-conscious, and what makes that person the same over time is having that same consciousness.

Locke's account does not constitute a complete view of personhood, because it raises all sorts of different questions concerning what self awareness or self consciousness consists in. One major objection that Joseph Butler raised to Locke's view of personal identity is that this account is a circular one.

"And one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes."[40]

Butler pointed out that self-consciousness and personhood go hand in hand, but that does not mean that this self-consciousness is what defines personhood. Consciousness of oneself cannot be a condition for being a person, because self-consciousness means that you are aware of your self, which presupposes that you are a self, and hence a person. To have self-consciousness you need to be a person, so you cannot say that what it means to be a person is to have self-consciousness. This is obviously

---

[40] Butler, J. 'Of Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 99-106

circular. This can also be illustrated when discussing self-conscious memory over time. Shoemaker argues that since we can have false memories, memory needs to be successful to constitute personal identity.[41] In other words, you might remember that you locked the door of the office before leaving it when you did not. Since these cases occur often, memory cannot be a reliable condition for personal identity over time. The way to mend this is to say that it is only true memories that constitute conditions for personal identity. So you are the same person as the one who locked the door of your office if and only if you have a memory of locking the door of your office, and you did in fact do that. But what makes something a true memory is the fact that it happened to you, in other words that you are the person who did it. But that is obviously circular, because one of the conditions for you to be the same person as the one who did something is that you are the same person as the one who did it. Memory cannot constitute personal identity, because if it is any memory whether true or false, then it becomes too liberal, allowing identity between things we would not want to consider identical. It would make you identical to the person who locked the door of your office, when no one actually did that. But if Locke says that it is only true memories that are conditions for personal identity, then that is circular and doesn't help us at all. The only way for self-consciousness to be considered a non-circular condition for personal identity is for it to be spelled out in a different way. So one must

---

[41] Shoemaker, S. 'Personal Identity and Memory' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 119 – 134

give an account for what it is to be self-conscious (in the way that is relevant to us) without appealing to personhood, and so avoiding circularity.

As I mentioned in the previous chapter, Flew[42] argues that Locke dug his own grave, because he distinguished his concept of person from anything else, leaving him with nothing but a circular account. This is a major objection to Locke's view, but I will argue that we can still talk about the forensic notion of a person by appealing to self-consciousness, as long as we give a non-circular account of what this self-consciousness consists in. I will start off by presenting an account of what a conscious state consists in, and then I will use that to give an account of how that relates to the kind of consciousness relevant for our present purposes.

b. Consciousness

Many have considered consciousness to be the mark of the mental. In other words, they believe that what makes a state a mental state is the fact that it is conscious. David Rosenthal proposes a different mark.[43] He believes that some mental states are conscious and others are not. There are many times when you are in a mental state but are not aware that you are in it, and so there must be an account of mental states that does not rely on consciousness. He argues that the definition of a mental states is whatever either has phenomenal properties or intentional content, and this makes it

---

[42] Flew, A. 'Locke and the Problem of Personal Identity' *The Journal of the Royal Institute of Philosophy*, vol. 2, pt. 6, pp. 53-68, January 1951

[43] Rosenthal, D. 'Two Concepts of Consciousness' from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 462-479

possible for some mental states to be conscious and others not. He argues for the view that a conscious mental state is one that is accompanied by a higher order thought about it, so if you are in a mental state, and have an accompanying thought about this mental state, then it is a conscious one. You might be in pain without realizing that you are in pain because you may, for example, be distracted and be thinking about other things. But when you have a thought about the pain you are in, then that pain is a conscious one. This is an account of conscious mental states that is non-circular. If you define a person as the type of being which is conscious of itself as itself, and then give an account of what this self-consciousness consist in, then you can move forward and not worry about having a question-begging theory. In other words, the definition of a conscious state includes an intentional state that has another mental state as part of its intentional content, which means that it is a thought about another thought, and since this is a distinct thought with its intentional object being something separate from it, then the circularity objection fails. Although this does not solve problems concerning consciousness of oneself, only consciousness of one's states, it clears up a lot of the confusion. What we need to do now is relate this theory of consciousness to our normative concerns and the question at hand.

c. Relevant Consciousness

Just like Locke, Harry Frankfurt[44] distinguishes between persons in the forensic sense of the term and human beings as they are classified

biologically. He believes that persons are the types of beings that have freedom of the will, and he tries to define what this will consists in. He believes that beings that are capable of having what he calls "second order desires" can be persons. In other words, if some beings are capable of having desires about other desires, then they are candidates for personhood. This is a certain version of consciousness of oneself as oneself. It is the view that people who are capable of viewing their mental states as their own and adopting attitudes towards them are somehow considering themselves as themselves. In fact, this is a more specific view that just says that as long as you have these types of mental states (namely the higher order ones), that makes you a candidate for personhood. However, Frankfurt adds that that is not enough for personhood, but what is a necessary and sufficient condition for personhood is for a being to have "second order volitions". Second order volitions are second order desires about desires that the agent wants to be effective. The necessary and sufficient condition for someone to be a person is that he is able to have a strong attitude concerning which of his desires he wants to be effective.

I will now turn to the argument that Frankfurt uses to establish this conclusion. I hope to show that Frankfurt's argument does not prove that higher order volitions are necessary and sufficient for personhood, but that higher order attitudes are. I will argue that what determines personhood has to do with moral character, and not with actions.

---

[44] Frankfurt, Harry 'Freedom of the Will and the Concept of a Person' from *The Importance of What We Care About* p. 11-25 (New York, Cambridge University Press, 1994)

Before turning to the arguments, one can see here the parallel to Rosenthal's theory of conscious states. Since Rosenthal is more concerned with the mark of the mental and consciousness as a subjective state, he is more interested in higher order thoughts than second order desires. But our concern here is a normative one. We are interested in agents that can perform actions and can be held responsible for these actions. So Rosenthal's account is helpful, but not the account we will rely on for the relevant consciousness we need. Frankfurt gives us an account similar to Rosenthal, but one more relevant to what we are interested in, actions and character. Frankfurt addresses the normative concerns that Rosenthal is not interested in. This account relates these second order mental states to actions, and more importantly to character.

Let us return to Frankfurt's argument. Frankfurt believes that persons are the class of beings that have freedom of the will. I said earlier that he is also concerned with Locke's notion of a person which is the type of being that is responsible for its actions. Frankfurt believes that one is responsible for his actions if and only if they are in line with his will, regardless of whether or not he can do otherwise[45]. So it is no surprise that he considers persons to be the only beings that have this freedom of the will. Since persons are the only types of beings that can have second order desires, and freedom of the will is requires having second order volitions, which are a form of second order desires, then only persons can have this freedom of the will. He writes:

---

[45] For more on this, see Frankfurt, H. 'Alternate Possibilities and Moral Responsibility' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p 1-10

"It is my view that one essential difference between persons and other creatures is to be found in the structure of a person's will… It seems to be peculiarly characteristic of humans, however, that they are able to form what I shall call "second-order desires" or "desires of the second order".[46]

He later adds:

'A person's will is free if and only if he is free to have the will he wants. This means that, with regard to any of his first-order desires, he is free either to make that desire his will or make some other first-order desire his will instead.[47]

There is a clear connection here between an agent's will and his actions, and this is what makes that agent responsible for this action. We might notice that so far, on Frankfurt's account, the only thing that is required for something to be a person is that it has first order desires that can motivate them to act. Of course that is not enough, because many non-human animals and non-persons act on their first order desires. Frankfurt adds that it is the effective first order desire that we are interested in because someone might have the desire, or even intention, of doing something without doing it. He also says that someone might have a desire *"but univocally want that desire to be unsatisfied"*[48]*.* So a person must have a desire that he wants to act on, which means that persons need something of the second order. However,

---

[46] Frankfurt, H. 'Freedom of the Will and the Concept of a Person' p. 12

[47] Ibid p. 24

[48] Ibid p. 14

second order desires are not enough to make something a person. A person is a being that has second order volitions. To illustrate the difference between second order desires and second order volitions he introduces the example of the physician who is researching drug addiction. For the purposes of his research, this physician decides that he would like to know what it feels like to be a drug addict.[49] However, he does not want to take drugs, so he desires to have the desire for drugs, but does not want that desire for drugs to be effective. This helps us distinguish between having a second order volition and a simple second order desire. Since his will is identified with the desire that he wants to be effective, and not the desire that he just wants to have, then that is what is needed for personhood. So in order for something to be a person, it must have second order volitions. It is not enough to have second order desires. He adds:

"It is logically possible, however unlikely, that there should be an agent with second-order desires but no volitions of the second order. Such a creature, in my view, would not be a person."[50]

He gives no argument for that, except for his argument in the previous section proving that higher order volitions are necessary and sufficient for something to be a person. Frankfurt's account relies heavily on the will and actions caused by this will, and that makes it very difficult for him to talk about personal identity over time. Personhood at a given moment for Frankfurt has

---

[49] Ibid p. 14-15

[50] Ibid. p. 16

to do with actions as much as responsibility does. It is true that what distinguishes persons from non-persons is the fact that they can consider themselves to be themselves and identify with some of their actions and not others, but that does not necessarily mean that they have to be decisive on which desires they would like to be effective. He does not give us an argument for why this makes these agents the type of beings that can be held morally responsible for their actions. All he says is that this defines the will and that beings with freedom of the will can be held responsible for their actions. In the next section, I will try to bridge the gap between what kinds of beings we can hold responsible for their actions and these beings' conceptions of right and wrong. It because of their higher order attitudes in general and their ability to reciprocate certain treatment (the condition mentioned above) that persons are able to make moral judgments, and that is what makes them the types of beings that can be held responsible for their actions. They are the types of beings that can understand what responsibility entails and can hold others responsible for their actions. This is what makes them members of the moral community. So we can say that personhood has to do primarily with character, and not actions. Actions only become necessary when we are discussing responsibility, as the next chapter will show.

4. Rationality and Conception of Right and Wrong

We have argued earlier that two conditions for being a person are sentience and rationality. Immanuel Kant[51] believes that being rational is a necessary and sufficient condition for having a conception of right and wrong. That is what makes something a moral agent. It is not within the scope of this work to discuss Kant's ethical theory, but this aspect of it is relevant for our discussion. I am not be committed to that strong a claim, but I do agree with Kant that we need an argument for why it is only persons and nothing else that can distinguish between right and wrong. For us to be able to hold an agent responsible for its actions this agent must be able to at least understand what responsibility is and so must be able to have a conception of right and wrong. If we had a being that acted but did not understand what responsibility meant, or what it was for an action to be an immoral one, then it seems counterintuitive to hold that being responsible for his actions. We may punish or reward them in an attempt to condition them not to do it again, but that hardly counts as holding them responsible. Responsibility, as I will argue in the next chapter, involves recognizing one another's intentional states, so for a being to be held responsible for his actions he must understand the attitude that others are taking towards his intentional states. That of course does not mean that he has to have well spelled out ethical theory, but that he can at least understand that some of his actions can be viewed as good or bad. It is intuitively unappealing to consider someone blameworthy of an action if he has no understanding of what it is for an action to be wrong. To be

---

[51] Kant, Immanuel *Groundwork for the Metaphysics of Morals* Tr. Thomas E. Hill Jr. and Arnulf Zweig (New York, Oxford University Press, 2002)

held responsible for his actions, a person must be a member of the moral community. This section discusses what it entails for one to be a member of the moral community. In the second section of the next chapter I will discuss why second order endorsements are necessary for holding responsible for a particular instance.

It is not only rationality that is required for having a conception of right and wrong. Having this conception seems to also be related to the condition concerning higher level attitudes that we discussed in the previous section. Persons must have some preferences concerning which desires they would like to act on, and they must be able to evaluate their desires. They must be able to recognize these intentional states and adopt certain attitudes towards them. Another condition that explains how persons can distinguish between right and wrong is the reciprocity condition. If one is able to take an evaluative attitude towards one's own intentional state, and can recognize intentional states in others, this enables one to have attitudes towards another's intentional states. That is how we are able to make moral judgments and evaluate other people's actions. When you see a man commit a terrible crime that is obviously planned, you recognize his desire to harm another, and you take an attitude towards that desire. If you see a man harm another by mistake you do not have the same attitude, because you do not recognize any desire of his that you take an attitude towards. Once we are able to recognize our own intentional states, other people's intentional states, and take attitudes towards both, we come to recognize the difference between

right and wrong. It is the existence of these states as well as our recognition of them that explain how we can hold a person responsible for his actions. This enables us to understand what responsibility means. A more detailed discussion of what kind of responsibility we are concerned with will come up in the next chapter, when I discuss the Strawsonian notion of responsibility.

Strawson asserts that a person can only ascribe mental states to himself if he is able to make similar sorts of ascriptions to others.[52] The way we ascribe states to others is through observation, but self-ascription is more immediate. In a different article Strawson discusses responsibility. He argues that it is through recognizing the intentions of others that we come to hold them responsible.[53] If we relate these two things to each other, by appealing to higher order attitudes, we realize that holding another responsible for an action involves taking an attitude towards the intentional state of that person. This might also help us better understand our intentional states and the way we come to evaluate them. This also explains our ability to distinguish between right and wrong. We think that an action is wrong when we have a negative evaluation of the desire that leads to it. We think an action is right when we have a positive evaluation of the desire that leads to it. When we recognize these desires in other people, we come to judge them in a way parallel to the way we judge our own desires. We come to see these agents

---

[52] Strawson, P.F. 'Persons' from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 104-116

[53] Strawson P. 'Freedom and Resentment' from *Free Will: Oxford Readings in Philosophy* by Gary Watson (ed.) (New York, Oxford University Press, 2003) p. 72-93

as persons just as we are, and we become capable of treating them as such. This is what makes us both members of the moral community in which we think of each other as agents and recognize one another's intentional states. We are somehow acting as judges of these intentional states, and we assess actions based on them, similar to the way that we judge our own intentional states and actions. It is not surprising that beings that adopt attitudes towards their own intentional states and can do the same with others are persons. So it is the existence of these attitudes or assessments of the higher order that make us persons. So something is a person in virtue of having higher order attitudes, which entail examining some intentional state and evaluating them. This grounds our ability to be responsible and to hold others responsible for their actions, something that is crucial for being persons.

The argument above suggests that we only need higher order attitudes to be persons, and not second-order volitions. That, in conjunction with the other conditions mentioned above, makes these the types of beings that can be held responsible for their actions. They are persons because they have a conception of ethics, and can distinguish between right and wrong actions. These are beings that hold others responsible for their actions, and can understand what it means for someone to hold them responsible for their actions.

5. Personhood without Action

Given our conditions for personhood, it seems unclear what one should say about cases in which there is a being who fulfills all of the above

conditions, but can never act. We started off by saying that we are using the term ‘person’ as a being to whom we can attribute responsibility. However, it might be a little tricky to claim that someone who is incapable of performing actions is not a person, because we can never attribute responsibility to him. In other words, this human being, who has higher level attitudes, and perhaps a conception of right and wrong, will never be held responsible for an action only because they will never be able to act. For instance, someone who is a mute quadriplegic since birth might not be able to perform many actions or bring about a variety of different state of affairs, and so can hardly be held responsible for many actions, or states of affairs. But the fact that they cannot perform many actions should not exclude them from personhood. In these cases, it might be helpful to think counterfactually about what they would have done had they been able to act. Now of course that will never warrant holding them responsible for actions that they had never committed nor will ever be able to commit. This is just to say that they should be considered persons in our sense of the term. In other words, even though we will not be able to hold them responsible for many actions, they can still be considered persons in the sense we are concerned with. They still fall under the category of the types of beings we can hold responsible for their actions. In other words, although they will rarely be responsible for any particular action, they are still considered persons.

One way to explain the difference between a person in general and an agent who is responsible for a particular action is to say that persons have

character. Character has nothing to do with a particular action or a set of actions, but it has to do with what type of intentional states someone approves or disapproves of. Someone can be a person even if he cannot act because he still has character, and condemns certain intentional states while approving of others. Many problems arise for Frankfurt because he entangles personhood with action when he does not need to. Personhood has to do with character, and the way we can make sense of someone's character is through his distinctive psychology. The distinctive psychology that is relevant for moral considerations has to do with the attitudes adopted towards intentional states, whether or not these intentional states are effective.

Perhaps a comparison here might be helpful. Consider a case where a being has higher level volitions that are never effective. Such a case is widely discussed in the literature. For instance we can consider the case of Nada, a drug addict who really does not want to be a drug addict anymore but cannot help succumb to her desires for drugs. Nada is a person, although her second order volitions are not effective. This case is logically identical to the case of the person who never acts in terms of personhood. Both of them have higher order attitudes. They recognize their intentional states and adopt attitudes towards them. They might be able to recognize intentional states in others and adopt attitudes towards them as well. The only difference between these two cases is that in the case of Nada there are actions to discuss, whereas in the case of the quadriplegic there are none. Of course that does not necessarily mean that we are justified in holding Nada responsible, we have

to see if the conditions of responsibility apply to his case or not. This is different from the mute quadriplegic because in the case of the latter, there are no actions concerning which we need to think of responsibility. So whether or not Nada is responsible for her action is a question that arises, but there is no parallel question for the mute quadriplegic. These two cases can be distinguished from cases of beings that have no higher order attitudes at all. This brings us to the next section, which is also widely discussed in the literature.

6. Wantons

Frankfurt introduces the term “wanton” to refer to human beings who are short of personhood because they do not have any second order volitions. In other words, although a wanton might be a rational being, and might even be aware of his first order desires, he does not have any preferences concerning which of these first order desires he ends up acting on. Frankfurt argues that wantons are not persons, because they have no second order volitions. Since they are just acting on these first order desires, they are no different from other animals. Although a wanton might be completely rational and calculative when seeking to satisfy his strongest first order desires, he is not a person. Persons are distinct from other creatures, according to Frankfurt, because of these second order volitions that other beings do not possess. Wantons lack that, so they are not persons.

Since I have a different account of personhood from Frankfurt, I will have to define wantons in a different way. On my account, a human being

that has no higher order attitudes at all is a wanton. In other words, if a being cannot adopt any attitude towards his first order desires, then that being is not a person. If they do have higher order attitudes but no higher order volitions, I still consider them to be persons.

One example that might be puzzling for Frankfurt which my account can deal with is the following. Consider Aaron, a man who decides that he would like to act on his strongest first order desire, whatever that first order desire may be. He has no specific higher order volition, but a general higher order volition concerning the way he will conduct himself. Note that we cannot say that Aaron has a higher order volition in line with every action he performs, because that would mean he wants to act on that particular desire. All he really wants to do is act on the desire that is the strongest, whichever one it is. His whole reasoning is that he does not want to identify with a particular desire, but he wants to act on desires that are strongest. It is not quite clear what Frankfurt would say about this man. He can say that Aaron is a wanton, because he does not have any specific higher order volitions. This does not seem right. Aaron thinks about his desires and the way he would like to conduct himself. He considers himself to be himself. Frankfurt might go the other way and say that Aaron is a person, and that he has higher order volitions in line with each one of the desires that he acts on. That seems to be wrong as well, because Aaron specifically wants this to be false of him. He does not want any particular desire to be effective, just any desire that happens to be the strongest one. The strongest desire might change from

one day to the next, and so we cannot say of him that he has a higher order volition in line with this desire.

On my simple higher order attitude account, we can deal with this case differently. Recall that on my account, if the basic conditions are satisfied, a man is a person if he has higher order attitudes of any kind. Aaron has a way of life, recognizes his desires, and adopts attitudes towards these desires. He evaluates them in a strange way, but he does evaluate them nonetheless. He can recognize intentional states in others and can take an attitude towards them as well. He should be able to tell the difference between right and wrong and understand what responsibility consists in. Although he does not have any specific higher order volitions, he is still a person because he has the relevant self-consciousness that we had discussed in an earlier section.

This account of wantons seems to be what we have in mind when we talk about them in ordinary discourse. When talking about wantons, we generally mean people who are amoral, or people who have no moral consideration or understanding. If someone has higher order attitudes (along with the other conditions) then that person is capable of stepping back and taking an attitude towards his own actions. That being is a person. Anything else is not.

Conclusion

In this chapter, I have presented several conditions which together constitute the necessary and sufficient for making something a person. The first and basic conditions were embodiment and sentience. Another basic

condition was that the beings in questions must be intentional systems. To be intentional, a system must possess rationality, have intentional states attributed to it and has to be the type of system that we can adopt the intentional stance towards. Another condition for personhood is the reciprocity condition, which states that a being has to be able to adopt the intentional stance and attribute intentional states to others. Candidates for personhood must also have linguistic abilities. The last condition for personhood is that these beings must have higher order attitudes. They should be able to recognize their own intentional states and adopt attitudes towards them. It is through adopting attitudes towards one's intentional states that one comes to have character. It is through these higher order attitudes as well as the ability to take the intentional stance that persons come to recognize the difference between right and wrong, and become capable of holding others responsible. This is what makes these beings the type that can be held responsible for their actions. Now we shall turn to what makes a person responsible for a particular action.

Chapter 3 Responsibility

In this chapter I will be deal with the following question: Given some person and some action, what must be his relation to that action to make him responsible for it? So I will be discussing the conditions that make a person responsible for an action.[54] I will start off by defining the kind of responsibility we are concerned with and then I will present the conditions that make one responsible for an action.

1. Kind of Responsibility in Discussion

Before diving into the discussion of what conditions make a person responsible for an action, it might be helpful to talk about what we mean when we discuss responsibility. People often discuss responsibilities they have towards others, or even towards themselves. These responsibilities are defined in terms of obligations and expectations of that person towards others. I will not be discussing this sort of responsibility, but I will be concerned, as I mentioned above, with the conditions that make someone responsible for an action.[55] We can distinguish here between two main questions that need to be addressed: First, when do we hold someone responsible for an action? And second, are we justified in holding this person responsible for that action? One answer to both those questions is presented

---

[54] I will not be discussing responsibility for state of affairs, even if the state of affairs was caused by the person's actions, because that will bring up a slew of questions that are not within the scope of this work.

[55] An account of this is also given by R.J. Wallace in *Responsibility and the Moral Sentiments* (Cambridge, Harvard University Press, 1994). Although both our theories have the same basis, we disagree on how to elaborate on Strawson's theory.

by P.F Strawson[56] and has been appropriately dubbed "the Strawsonian view of moral responsibility"[57], and I will be adopting a slightly modified version of this account.

a. The Strawsonian View of Moral Responsibility

Strawson argues that it is a result of our social practices that we come to hold people responsible for their actions. Moral responsibility has nothing to do with metaphysics. We happen to be the sorts of creatures that act and react to each other's actions in certain ways that make us hold one another responsible for these actions. So the truth or falsity of the thesis of causal determinism will not shed any light on the issue of moral responsibility. The way we define moral responsibility on this view is naturalistic, and involves describing how it is that certain reactions come about, and when these reactions are considered justified. So it is natural for us to hold people responsible for their actions, and sometimes justifiably so. On this view someone is responsible for an action when he is the appropriate recipient of certain types of reactive attitudes. In other words, a person is responsible for an action when it is reasonable for you to have a certain type of attitude towards them regarding the action that they had preformed. Such reactions can include praise or blame, or even indignation.[58] Of course the specific

---

[56] Strawson, P.F. 'Freedom and Resentment' from *Free Will: Oxford Readings in Philosophy* by Gary Watson (ed) (Oxford, Oxford University Press, 1982) p.72-93

[57] John Martin Fischer and Mark Ravizza *Responsibility and Control* (New York, Cambridge University Press, 1998) p. 5-8

[58] R. J. Wallace describes responsibility in terms of three emotions, and he gives an account based on expectations people have of one another. His is one attempt at filling in a fuller normative story to

reaction one has depends on the moral worth of the action in discussion. In order to understand what this all means, let us turn to a more detailed discussion of this view.

It is controversial whether Strawson's real goal was to identify conditions that make one responsible for an action. Some believe that his view is only descriptive and serves only to explain when we hold someone responsible for his actions. Others believe that it has normative elements in it, and although it is not a comprehensive normative view, it does aim at laying the conditions of when a person is responsible or not. I'm going to use the first interpretation, not because I believe there is more textual evidence in Strawson's work supporting it, but because it is the one that I'm interested in for the purposes of this project. Whether or not it was his real concern is not of crucial importance, but it is of crucial importance for me, and I will try to use his account to that effect.

As mentioned above, at the core of this view is a naturalistic descriptive account of responsibility. Through describing our reactions to one another's actions, we come to understand what it means to hold one responsible for one's actions and when this reaction of ours to someone's action is a reasonable one, then that person is considered responsible for that action. So ascription of responsibility is a result of our social interactions, since we expect a certain degree of good will from one another. When

---

Strawson's account. However, I do not believe his story to be successful for it fails to explain many attitudes that we do have that do not involve expectations at all. Nor is he ultimately successful in relating why persons are the only kind of beings that can be held responsible for their actions. For his full account, see R. J. Wallace, *Responsibility and the Moral Sentiments* (Cambridge, Harvard University Press, 1994)

someone does not meet this expectation, we cannot help (and perhaps reasonably so) but feel resentment towards that person. Having certain attitudes and reactions is an essential and vital part of our personal interactions and we cannot just rid ourselves of it because we have uncovered some metaphysical fact, for example the causation of the action. It is almost impossible for people to quit having these attitudes towards others just because of something a metaphysician has told them.

We must note that there is a difference between regarding one as an object of social policy or to be cured and treating them as a fellow member of the moral community. Strawson distinguishes between what he calls taking an objective attitude and a participant one towards other people.

"To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided, though this gerundive is not peculiar to cases of objectivity of attitude."[59]

We can even say that taking the objective attitude towards a person means objectifying them in some sense. Of course this does not necessarily have to be degrading objectification, because some doctors objectify their patients in order to cure them and figure out how to help them. It might very difficult for a doctor (or even legislator) to consider all of their patients (or

---

[59] Strawson, P. 'Freedom and Resentment' p. 79

subjects) as persons with goals, desires, and needs, so they need to adopt this objective attitude in order to be able to deal with the questions at hand. So they often dissociate themselves from the patient (or subject) and treat them as they would treat an object that needs to be fixed. On the other hand, when one makes a new friend or falls in love, there is a different attitude that he takes towards that person. This is a participant attitude.

"What I have called the participant reactive attitudes are essentially natural human reactions to the good or ill will or indifference of others towards us, as displayed in their attitudes and actions."[60]

When holding people responsible for their actions, we are engaged in forming participant reactive attitudes. In some sense, taking the participant attitude toward someone involves regarding them as a fellow member of the moral community who has desires and goals of their own. We might use the terminology from the previous chapter by saying that participant reactive attitudes involve taking the intentional stance towards someone and adopting attitudes towards their intentional states. This is what treating them as persons entails which the objective attitude does not involve. Higher order attitudes are not necessary for taking the objective attitude. Most of our interactions with other people involve taking the participant attitude, because when you are involved in a conversation with another person[61] or when you

---

[60] Ibid p. 80

[61] One interesting connection made between responsible desires and beliefs is made by Philip Petit and Michael Smith, who argue in a way similar to the one I argued in the previous section concerning responsible agents and make a parallel argument for responsible interlocutors and believers. To be able to have conversations concerning beliefs and desires, you have to assume that your interlocutor is a

are having a fight with someone, you are taking the participant attitude toward that person. So when we hold someone responsible for his actions, this involves taking a participant attitude towards that person, and not an objective one. When you hold someone responsible for his actions, you attribute certain intentions to them and you react to these intentions and the actions that result from them. It is normal for us to feel resentful towards someone who has intentionally harmed us and grateful towards someone who has intentionally helped us. It is these natural reactions that are involved in holding one responsible for his actions. Strawson gives an illustrative example.[62] If someone accidentally steps on your hand while trying to help you, we do not consider it reasonable for you to feel resentment towards that person, because after all, his desire was to help you. But if someone stepped on your hand in a malicious attempt to hurt you, we consider your feeling of resentment towards that person to be a reasonable one. It is the fact that his action was intentional that makes you resent him.

It is now clear how this ties up with our discussion in the previous chapter. When someone steps on your hand intentionally, you recognize their desire to harm you and you take an attitude toward this desire and that is why you naturally blame that person for hurting you. You have a natural reaction of resentment towards that person with respect to the act that they have

---

free thinker and agent, otherwise there is no point in the discussion. Being a free thinker involves having self-reflecting attitude towards your beliefs, and being a free agent involves self-criticism of one's desires. For their discussion on the subject see Petit, Philip and Michael Smith 'Freedom in Belief and Desire' *Journal of Philosophy* Vol. 3 No. 9 (Sep. 1996) p. 429-449.

[62] Ibid p. 76

done. You do not approve of that desire but consider it to be blameworthy in some sense. This is a purely descriptive account of what happens when you hold someone responsible for his actions.

b. Justified Reactive Attitudes

So far, we have described what the practice of holding one responsible for his actions involves. The next question that should be addressed is the following: Are we justified in holding this person responsible for this act? After all, this is the question that interests us in this project. Since we do not want to give a metaphysical answer or an answer outside the scope of the practice itself, we are left with one of two types of answers. First, we might try to justify these attitudes by appealing to the value of this practice, so we can say that we are never really responsible for our actions, but we should continue with the practice of holding one another responsible for our actions because of the value of this practice. So it is valuable for us to go on with this practice even if it involves self-deception. Such a line is defended by Susan Wolf[63]. Wolf gives us some interesting thought experiments that justify her position, namely that a world without participant reactive attitudes would not be one that we would prefer over our current world. Although her thought experiment brings to light the intuitive appeal of maintaining this practice, it is not a very strong argument in favor of the practice. Many people are sympathetic with the view that they would rather live in a less valuable world but be honest with themselves than live in self-deception just because it makes them feel better.

---

[63] Wolf, Susan 'The Importance of Free Will' from *Perspectives on Moral Responsibility* ed. By J.M. Fischer and Mark Ravizza (New York, Cornell University Press, 1993) p. 101-118

After all, why should we bother with philosophy if we are going to ignore its conclusions?

The second type of answer involves discussing nothing outside of the practice at all and giving conditions that make these attitudes justifiable within the context of this practice. This is a more appealing line which Strawson himself adopts. Although Strawson describes the practice of holding others responsible for their actions in an impressive way, he never specifies the conditions which make it reasonable to have these reactive attitudes. Whether or not this was his intention is besides the point, but he certainly does not give us such an account. If one were to read such an account in Strawson, he would have to fill in the blanks concerning what reactive attitudes can be considered reasonable and justifiable, which is precisely my project. All Strawson does in terms of justifying some participant reactive attitudes is to give excusing conditions. So instead of telling us when one is reasonably held responsible for his action, he tells us when it is unreasonable to hold one responsible for his actions. He spells out these two vague excusing conditions. The first excusing condition specifies that there are certain conditions in a situation that exempt the agent from responsibility. The second excusing condition is that the being who committed the act was not a moral agent, or that conditions of personhood do not apply to that being.

The first chapter discussed what conditions are constitutive of a person. So the second set of excusing conditions was addressed. If one fails to satisfy one of the conditions necessary for personhood then the second

excusing condition applies. It is not reasonable to have participant reactive attitudes towards non-persons, or non-agents, since reactive attitudes are only reasonable when held between two members of the moral community.

This chapter will be about the first set of excusing conditions. So if an agent has performed an action but is not responsible for that action, we need to spell out the conditions that exempt him from responsibility. Strawson considers it obvious which conditions should warrant exemption of responsibility. Unfortunately this is not the case, since there are too many internal and external conditions for us to just give an answer that relies purely on common sense. My objection to Strawson is the he does not do enough to spell out the two excusing conditions which I believe that he does not do enough to spell out. Since, as I mentioned above, I dealt with one of these conditions in the previous chapter, I will turn now to specifying the conditions that would prevent us from reasonably holding an agent responsible for his action. But before turning to that, there are a few distinctions that help clarify the discussion.

c. Character and Responsibility

First, one must draw a distinction here between the attitude we might have towards a person, and the attitude we might have towards a person with respect to an action they performed. We might find someone to be appalling and perhaps even morally abhorrent even if we do not hold them responsible for any action at all. Take for example a case we discussed in the previous section, we can find a certain person (the mute quadriplegic) to be morally

appalling and maybe even the appropriate recipient of some attitudes purely based on their character, and the types of higher order attitudes they have. This should be distinguished, however, from holding them responsible for an action.

A person's actions might be in line with his character, but he might also be acting on a desire that is alien to him, and thus adopt a negative higher order attitude toward his action. Such a person will not be responsible for his action. A person might also have a positive attitude towards another person's action that we disapprove of, but we cannot hold him responsible for that action. For instance, Fadi may have a positive higher order attitude towards the rape that Omar had committed, but we cannot hold Fadi responsible for this rape. We cannot have a justified participant reactive attitude towards one unless he has done the action, so we cannot have a justified participant reactive attitude towards Fadi, but we can towards Omar, at least with respect to this action. What we are concerned with is whether they are the justified recipients of reactive attitudes in relation to those actions. In other words, their relation towards the action is what the responsibility for that action consists in, and the attitude we attach to it has to do with what we consider the moral worth of an action to be.

d. Responsibility and Moral Worth of Actions

There is another distinction that we should keep in mind, and that is the one between holding someone responsible for an action and the moral worth of the action under consideration. There is a difference between saying

that one is the justified recipient of some reactive attitude with respect to some action and specifying our attitude toward the action. Two people may agree that they hold a third person responsible for his actions, but might disagree about whether that person should be praised or blamed. For instance, Lama and Ameen are both judging Hassan who has killed someone because that man had killed his father. If Lama finds this action heroic and virtuous and Ameen finds it vindictive and vicious, Lama and Ameen nonetheless agree that Hassan is responsible, since they both think it justifiable to have some reactive attitudes. They both believe that it is reasonable to have some reactive attitude towards that person, but do not agree which particular attitude should be taken. Since we are not concerned here with the moral worth of actions, the focus will be on whether an agent is the justified recipient of some reactive attitudes with respect to an action, whatever these reactive attitudes might be.

So our discussion of responsibility for an action will revolve around the question of when it is justifiable to have certain reactive attitudes toward an action and when it is not. This will be what I will be considering when discussing moral responsibility for an action. Despite the fact that this is tied to the questions concerning reward and punishment, we must draw the distinction between the issue of moral responsibility and that of reward and punishment, and this is the topic of the next section.

e. Responsibility vs. Punishment/Reward

Many have argued that a person is responsible for an action when he is the one that should be punished or rewarded for it.[64] This view is too simplistic. It is important to distinguish between whether we are justified in rewarding or punishing a person and their being responsible for their actions. Although some have argued that responsibility only involves bestowing rewards and punishments upon agents, there is a difference between the two. Others have argued that people are never responsible for their actions, but we ought to punish people in order to rectify their behavior. We can think of this as conditioning people into acting the way we want them to act.[65] But that is not what we are discussing when we talk about responsibility. The section above should have illustrated that the ascription of responsibility only involves participant reactive attitudes, and might have nothing to do with rewards or punishments. Such examples are often found with spouses, for instance Najla may blame Maan for forgetting her birthday, or praise him for remembering every single detail about the night he proposed to her, without thinking about or invoking any kind of reward or punishment. Such is the nature or personal relationships, since they involve participant reactive attitudes but need not involve rewards and punishments. We are therefore able to distinguish between holding someone responsible for his actions and believing that they ought to be rewarded or punished. In some cases we might recognize that

---

[64] Schlick, Moritz "When Is a Man Responsible?" in *Problems of Ethics* tr. David Rynn. (New York, Prentice Hall, 1939) pp. 143-156

[65] Bentham, Jeremy 'Punishment and Deterrence' from *Principled Sentencing* ed. Andrew von Hirsch and Andrew Ashworth (Edinburgh, Edinburgh University Press, 1992) p. 62-66

someone is responsible for an act that they had committed but still not punish/reward them. In other cases, the opposite might be true. Cases of this sort are often found when attempting to give children a moral education, since we reward and punish them for their actions without holding responsible for these actions. Whether one is justified in these cases would of course depend on his theory of punishment (or reward). People who believe in retributive punishment consider that punishment should be given to a person if and only if he is responsible for the action.[66] On the other hand, those who adopt a consequentialist account do not agree, because they maintain that there are cases in which the consequences of not punishing a responsible agent are better than those of punishing him.[67] For instance, if Will Ferrell was caught shoplifting, and you knew that arresting him or fining him may cause all kinds of riots, then some consequentialists would argue that it may be better to exempt him from punishment in order to avoid the unfavorable results of punishing him. We must note, however, that although a discussion of moral responsibility will not solve all questions concerning punishment and reward, it does shed light on those issues. But that is not within the scope of this presentation.

Since all these distinctions have been clarified, it is now easier for us to understand the topic under discussion. We now know what kind of responsibility we are discussing, and how it is to be distinguished from many

---

[66] Kant, Immanuel 'Justice and Punishment' from *Philosophical Perspectives on Punishment* ed. Gertrude Ezorsky (New York, State University of New York Press, 1972)

[67] Bentham, Jeremy, 'Punishment and Deterrence' p. 66

of the issues that it is usually confused with. We will now turn to Strawson's first set of excusing conditions. The next section will deal with the following question: Given a moral agent and a particular action that he has done, what conditions need to apply for him to be considered a justified recipient of some participant reactive attitude? Let us look at the proposed answer to this question.

2. Autonomy vs. Heteronomy of the Will

In the Groundwork for the Metaphysics of Morals[68], Kant famously distinguished between autonomy and heteronomy of the will. He believed that what makes one responsible for an action is not that was in accord with his will, but that his action was caused by his will. He called the former heteronomy and the latter autonomy of the will. If one were to put it in my terminology, he would assert that one is not responsible for an action unless it was caused by his higher order attitudes. So if a person acted on a desire that he could not help, but this desire was one that he had a positive attitude towards, then he would not be responsible for this action.

I will argue against this, maintaining that the causal role of this attitude is of no relevance in the determination of responsibility. What is crucial for moral responsibility for an action is the attitude the person has towards that action. To put it in Strawsonian terms, a person is not excused for his action only because it was not caused by his will.

a. Causal Role of Desires

---

[68] Kant, Immanuel *Groundwork for the Metaphysics of Morals* Tr. Thomas E. Hill Jr. and Arnulf Zweig (New York, Oxford University Press, 2002)

In order to understand what it means for one to be responsible for his action, we need to have an understanding of what constitutes an action. Following Donald Davidson, I will take it that:

"If an event is an action, then … under some description(s) it is intentional."[69]

As mentioned in the first chapter, our interest in this work is the intentional description of these events. But we need to understand what it means for something to be intentional, or to be done with an intention. To this Davidson answers:

"…an action is performed with a certain intention if it is caused in the right way by attitudes and beliefs that rationalize it."[70]

The further question concerns what kinds of attitudes these are. These pro attitudes include *"desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind."*[71]

One may also inquire as to what Davidson means by being caused "*in the right way"*. I am not concerned with giving a detailed account of what constitutes an action. For the purposes of this work I am concerned with the types of beings that act and the kinds of actions we should be held responsible for. Davidson's account tells us that actions are events caused by

---

[69] Davidson, Donald 'Agency' from *Essays on Actions and Events* (New York, Oxford University Press, 1980) p. 61

[70] Ibid p. 87

[71] Davidson, Donald from 'Actions, Reasons and Causes' from *Essays on Actions and Events* (New York, Oxford University Press, 1980) p. 4

desires (and other pro attitudes). I do not need to discuss the particular way in which these desires cause the action since all that I require is that they be caused by some desire or pro-attitude, and that is simply because it is a condition of action in general. It does not follow from this that the connection between higher order attitudes and the action must be causal for the agent to be responsible for the action. One would need a further the argument for this Kantian view. I will argue in a later section that the only causal connection that is needed for one to be held responsible for his action is the one between the desire and the action. All that is necessary at the higher order level is having a certain attitude towards these desires.

One may object to this by saying that a pro attitude might be a second order desire, and that actions caused by these are the ones we ought to be held responsible for. I agree that an action might be caused by a second-order desire, for instance Amy might want to stop being a smoker and cause herself to refuse a cigarette when offered, although she has a very strong desire to take that cigarette. This may happen in one of two ways: either the second order desire directly causes the action or the second order desire causes her to have a first order desire that leads to her action. It might be that the second order desire caused the action in this case, but later on when she develops a desire to refrain from smoking (or lacks a desire to smoke) a first order desire causes her action. Amy is responsible both on Kant's view and mine. However, if she had a first order desire to refrain from smoking caused by nothing of the second order, and that caused her to refuse a cigarette,

then this will not necessarily exempt her from responsibility. This is a case in which the agent acted, but there was nothing of the second order in the causal chain. However, I will argue that even in this case she may still be responsible for her action because although it is not necessary to have the higher order attitude in this causal chain, it is sufficient have that higher order attitude. In the next section I will argue why it is important to have some attitude of the second order if one is to be held responsible for his action.

The example above helps us understand the distinction between something being an action and an action for which an agent is responsible. Agents are not responsible for all their actions, for instance, one might be coerced into doing something immoral like stealing, and although the person stole intentionally, we might not hold them responsible for stealing. A more detailed discussion of coercion will follow later on in this chapter. One might also act without realizing what they are doing, and we might want to exempt them from responsibility. For example, if Samia ignorantly pours a cup of poisonous coffee for her guest, we would not consider it reasonable to resent Samia and blame her for murdering her guest. After all, she did not intend to harm anyone, she was just trying to be a good hostess.

David Velleman raises the following objection to Davidson's story:

"In this [Davidsonian] story, reasons cause an intention, and an intention causes bodily movements, but nobody - that is, no person – does anything. Psychological and physiological events take place inside a person, but the

person serves merely as the arena for these events: he takes no active part."[72]

I am somewhat sympathetic with this objection. However, I do not think it is a reason to reject Davidson's analysis. One way to work this out is to keep the widely accepted Davidsonian account of action but somehow connect this account with some account of personhood and agency. This is what I am trying to do. My account defines persons in terms of higher order attitudes, and these are what make the agent responsible for his actions. In other words, the same thing that makes someone an agent also grounds responsibility. These higher order attitudes also make the agent a particular person. The distinctive psychology of a person includes the very thing that makes him responsible for certain of his actions, namely the higher order attitudes. The agent is not merely an arena where these events happen to take place, but these higher order attitudes are what make him a person and define his character.

Before describing what particular attitude an agent must have towards his desire to make him responsible for an action, we should note why the causal role of this attitude is of no critical importance in the determination of responsibility. One can think of two cases, in both of which actions are done by agents who have the required attitude towards their actions, but one of whom acted because of a desire he could not help but have. Consider, for

---

[72] Velleman, J. David (1992). "What Happens When Someone Acts?" in *Mind*, New Series, Vol. 101, No. 403, pp. 461-481.

instance two drug addicts, one of whom is more addicted to drugs than the other, but both have positive higher order attitudes to their habits of drug addiction. Now Ellen, the first drug addict, acted out of a desire that she cannot help but have, but she happened to have a higher order attitude of approval for that desire. Rachel, on the other hand, had a higher order attitude of approval for the desire to take drugs, and this attitude is what caused her to act, not the lower level desire. It is important to keep in mind that in this case, the attitudes of the two agents, Ellen and Rachel, are the same, and it is only the causal story that is different. Rachel's action was caused by a desire that was itself caused by the higher order attitude while Ellen's action was caused by a lower level desire in line with her higher order attitude. There is no difference with respect to responsibility between the two cases, and that both Ellen and Rachel should be considered blameworthy for their actions if we disapprove of them. What is crucial for moral responsibility for an action is the higher order attitude that the person has towards his action: whether they identify with the desire that caused it or not. Of course the causal story will play some part, because after all we are discussing actions, whose causes must be desires of the agent's. However, what makes an agent responsible for an action is that his desire is one which he identifies with, that is has a positive higher order attitude towards. Whether or not that action was in fact caused by this higher order attitude is irrelevant to moral responsibility.

b. Importance of the Higher Order

In the previous chapter I discussed why higher order attitudes are necessary for one to be a person, or a member of the moral community. Now we turn our attention to why higher order attitudes are necessary (and sufficient) for one to be held responsible for his action. Thomas Nagel[73] explains that when discussing responsibility there are usually two players, one who is the judge and the other who is the defendant. The judge places himself in the agent's position and then looks at the circumstances as they presented themselves to the agent. It is on the basis of such considerations that the judge decides whether he approves or disapproves of the agent's actions, and this is what our reactive attitudes are based on. I wish to take this a step further and say that the judge takes an attitude towards the agent's desires. I say that because as the Strawson example I mentioned above suggests, it is only after we recognize the agent's intentions that it becomes reasonable for us to hold our reactive attitudes. It is unreasonable for you to be resentful of your friend who stepped on your hand while trying to help you. It is, however, completely reasonable when you recognize that the man who stepped on your hand did it because he wanted to harm you. This suggests that it is this intention, or desire, that we have the reactive attitude towards. If the intention we attributed to the agent was not present then it becomes unreasonable for us to blame him for the action.

---

[73] Nagel, Thomas, *The View From Nowhere* (New York, Oxford University Press, 1986) p. 120-121

So not only does this higher order attitude make you part of the moral community, but it is also that attitude that makes you responsible for your action. It is his approval of the intention to harm that makes a man responsible for his action. In this case, the agent is capable of making himself the judge, and looking on the action the same way any other judge would. This is a way of acknowledging responsibility. When the agent has a positive attitude towards what he is doing, then he is placing himself in the judge's position and approving of his own action. In this the agent comes to accept responsibility for his action, for he recognizes himself as the agent and approves of himself while he is acting. It is this higher order attitude that makes the agent not only a member of the moral community, but also one who accepts responsibility for his action. If he were someone else he would have had some reactive attitude toward himself. So having this higher order attitude seems to be necessary for one's being responsible for his own actions.

c. Second-Order Volitions

As mentioned in the previous chapter, Frankfurt focuses on higher order volitions, and believes that an agent is responsible for an action if and only if he has a higher order volition in line with it (although he states that this higher order volition doesn't have to be the cause of the action[74]). In other words, he argues that one is responsible for an action if and only if he acted on a desire that he wanted to be effective.

---

[74] Frankfurt 'Freedom of the Will and the Concept of a Person' p. 25

In the earlier chapter I discussed what it means to be a person for Frankfurt. We looked at his argument in support of the theory that higher order volitions are necessary and sufficient for personhood. Frankfurt argues that moral responsibility does not require alternate possibilities.[75] He later gives an account of freedom of the will that can ground responsibility regardless of whether the agent could have acted otherwise[76]. Given our notion of moral responsibility, it seems natural to adopt the Frankfurtian line. Frankfurt believes that what makes someone morally responsible for his action is the fact that it is in accord with his second order volitions. Since I argued in the earlier section that the higher order attitude is necessary for holding one responsible for an action, it follows that I should have an account of what makes someone responsible for an action that is similar to Frankfurt's. He argues that since second order volitions are necessary and sufficient for personhood, it is when acting in accord with these volitions that a person is acting with freedom of the will. The will is constituted in one or more desires that the agent has a second order volition towards. This means that the agent identifies with the desire that led to the action, and that this action is somehow expresses the agent's character. So since he is adopting a positive attitude towards it, it flows from the self and that makes it his own. This makes the agent responsible for this action.

---

[75] Frankfurt, H. 'Alternate Possibilities and Moral Responsibility' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p 1-10

[76] Frankfurt 'Freedom of the Will and the Concept of a Person'

I will adopt an account of responsibility that hinges on the higher order, but one that is slightly different from that of Frankfurt. I believe there are some objections to Frankfurt that he cannot meet, and this is the topic of the next section.

d. Some Objections to Second-Order Volitions

Before getting to my disagreement with Frankfurt, I would like to stop here to make a few distinctions that are helpful for my purposes, and to make sure we are not confusing our terminology. A higher order attitude is an attitude one holds towards a desire that he recognizes to be his own. This attitude may be a negative one or a positive one. If the attitude one has towards a desire that led to an action is a positive one then, as I will argue later on, he is responsible for his action. That should be distinguished from a second order volition, since the second order volition is a decisive commitment to want one of the desires to be the effective one. Having a positive higher order attitude is thus different from having a second order volition. Although a second order volition is to be distinguished from a higher order attitude, it is a special kind of higher order attitude, namely a positive and decisive one about what desire the agent wants to be effective. A second order desire, however, is simply a desire whose intentional content includes another desire. This is also different from higher order volitions (the argument for that is presented by Frankfurt and I will mention it later on in this section), and is clearly different from a higher order attitude, which is not another desire.

Let us now turn to my disagreement with Frankfurt. The reason that I disagree with Frankfurt is that on his theory, if one were to act on a desire that he did not have a second order volition in line with, but acted on it anyway, he would not be responsible. This is a case in which one has second order volition against an action, but did it anyway (whether or not he could help it). An example of this would be Martin, a married man who wants to desire sexual encounters with different women because he wants to be manly like his friends. However, he doesn't want to cheat on his wife, so he doesn't want this desire to be effective. So Martin does not have a higher order volition in line with cheating, but he might succumb to the desire he wanted to have and cheat anyway. Here is a case in which the agent had no second order volition, but still acted on a desire that he wanted to have It seems to be wrong not to consider Martin responsible for the action in this case. We still want to hold some people responsible for an action even though they didn't want the desire that led to it to be effective. Sometimes an agent commits an act despite not having a second order volition favoring it. After all, if someone knew that they were going to regret doing something (not because it was unethical, but because they just didn't want to do it) it doesn't mean that they are no longer responsible for it. Similarly, if someone were to perform an action, knowing it was wrong and that they ought not to do it, it does not mean that they were not responsible for it. According to Frankfurt, they wouldn't be responsible.

An example here might help illustrate this point. Frankfurt discusses the case of a doctor who is researching drug addiction, and for the purposes of his research, decides that he would like to know what it feels like to be a drug addict.[77] So he desires to have the desire for drugs, but does not want that desire for drugs to be effective. This is where Frankfurt's example stops, and he uses it to illustrate what a second order volition is, distinguishing it from a second order desire. I wish to extend this example a little further. In order for the doctor to really understand what it is like to be a drug addict, not only would he like to have this desire for heroin, but he would also want it to be part of his psychology. In other words, having a detached individual desire for heroin without affecting the rest of his psychology will not help him understand what it is like to be a drug addict, but it is when this desire is in fact his own that he comes to know what it feels like to be an addict. But although this desire is his own, he still does not want it to be effective. Now let us go on with the narrative and say that the doctor had that desire, and the desire became so strong he could not help but take heroin. Frankfurt here would argue that the doctor is not responsible for his action, because his second order volition was not in line with the desire that led him to act. I would argue that the doctor is responsible for the action if and only if he had a positive higher order attitude to the desire, but not a second order volition. It is not a sufficient reason to excuse the doctor just because he did not have a second order volition towards the desire that he desired to have. In fact, it

---

[77] Ibid p. 14-15

seems a little bizarre not to hold him responsible for acting on a desire that he endorses. This is a point on which I take issue with Frankfurt.

It might be that the same intuition driving my objection to Frankfurt is one that is explicitly stated by Gary Watson[78]. Watson poses the question:

"Can't one be a wanton, so to speak, with respect to one's second-order desires and volitions?"

Watson argues that since second order desires are themselves desires, then they might be in conflict with one another. The agent might not have a preference concerning which one of these second order desires wins out. Another way of stating this objection is to ask what the special status of these second order volitions is. Frankfurt emphasizes the decisiveness of these second-order volitions. Since the agent has a preference concerning which of these desires he would like to win out then he believes Watson's objection to be moot. However, Watson responds to this by pointing out that this decisiveness might be one that happens at the first order, and so one can just be motivated by one desire rather than another. There is no role that these second order volitions play that adds anything to the agent.

Watson's objection raises the following questions: why the second order? Frankfurt believes that the answer has to do with volitions, and the decisiveness of these volitions. Some might argue that this answer is circular, because we are trying to define what a person is, and it cannot just be that the person is what he wants to do. My answer is quite different. In previous

---

[78] Watson, G. 'Free Agency' Journal of Philosophy, vol. 7, pt. 2, pp. 205-220, April 1975

sections I have argued for the importance of the second order, without mentioning decisiveness or volitions. It is not the decisiveness that makes one responsible for his actions, but recognizing his desires and taking a particular attitude towards them. In doing so, the agent enters the moral community; in doing so he regards himself as a responsible agent, as he regards others as responsible agents; thus the person becomes responsible for his actions. In other words, when he takes an attitude towards his own desires, he adopts the position of the judge of his own actions, and becomes capable of being both the object and subject of attributions of responsibility. Recall in the previous chapter, we discussed why agents with higher order attitudes are the only types of beings that can understand responsibility. It is because in holding someone responsible for an action, you are adopting an attitude towards their desires and judging them accordingly. Once you are capable of playing this game, you automatically become an appropriate object of these kinds of attitudes. So being responsible for an action has nothing to do with volitions, but with another attitude that one has concerning his first order desires. To get back to Watson's objection, if there is a conflict of second order desires, and the agent does not care which of them wins out, he might still be responsible for his action. For example, consider Ghenwa, who wants to be both popular and benevolent, and she has no preferences concerning which of these two virtues she has. So one day, she is invited to a party that will be very enjoyable and people attending it will get noticed, but at the same time there is some charitable event that she would love to be part

of. These two are in conflict, and she has a desire to go to the benefit as well as a desire to go to the party, and she has positive higher order attitudes towards both. She does not particularly care which of these desires she acts on, as long as she goes to one of the two events. Surely in this case we will consider her responsible for either going to the benefit or not going, although she had no strong commitments towards going or refraining from going. Here is a case in which the agent has a conflict at the higher order level, but the fact that there was is a conflict does not exempt her from responsibility. As long as she recognizes these desires and has the appropriate positive attitudes towards them then she will be held responsible for them. This is because responsibility has nothing to do with decisive or indecisive volitions. The particular attitude that we need for responsibility is the subject of the next section.

e. Higher Order Endorsements

For Frankfurt, what makes someone a person is that he is a being with second order volitions. When he acts in line with these volitions then he is acting freely and hence responsibly. So in a sense, his actions are in line with what makes him the type of being that can be held responsible for his actions. There is an obvious and very strong connection between being a person and being responsible for one's actions.

Now on my account, someone is a person if he has higher order attitudes of any kind and not necessarily second order volitions. I have argued that these higher order attitudes are what make someone a person, so what I

now need to do is give the specific type of attitude one has to have towards an action to make him responsible for it. In other words, what I need to do now is bridge the gap between what makes someone a person, that is a being that can be held responsible for its actions, and what makes him actually responsible for a particular action. What makes someone a person is that he has higher order attitudes towards his desires, and what makes him responsible for a particular action is that he has a positive higher order attitude towards it.

I have defended a Strawsonian view of responsibility, namely that responsibility is embedded in our social practices. So given these social practices, I now need to give an account for what conditions make it reasonable to have some reactive attitudes towards a person with respect to an action. So on our journey we have gone from describing the types of beings that are engaged in this practice to grounding these concepts. Now we will turn to the last step.

So the question at hand is the following: what attitude should one have towards an action (or the desire that led to that action) that justifies being held responsible for it? The answer is that one has to have a positive higher order attitude of, that is approval or identification with, the desire. But the problem lies in trying to define what this approval or identification consists in. Both of these are vague terms that need some elaboration. One answer is that identifying with a desire means that one considers it to be his own, in other words not alien to him. Of course this seems to be a completely circular

account, but an explanation can mend that. One is a person if and only if he satisfies all the conditions mentioned in Chapter 2, including the last one which is having higher order attitudes of any kind. But what makes him responsible for an action is that the desire that caused this action was his own or not alien to him. In other words, if one recognizes the desire that led to his action and identifies this desire to be his own, then he is responsible for the action. Once he has identified this desire to be his own, he has made it part of his distinctive psychology, and this makes it part of his identity. This desire becomes part of him, and this any action that results from it his own action. The desire interacts with other desires of the agents, and becomes part of his distinctive psychology and it plays a part in forming his character. In this case, we might call this identification a “higher order endorsement”. The agent recognizes the desire, endorses it, and does not consider it alien to him. This phrase is short for that lengthy statement. One is responsible for an action if and only if he has a higher order endorsement of his action.

Perhaps an example might help us understand these abstract terms. Let us consider the case of the doctor from the previous section. On Frankfurt’s account, the doctor did not act in line with his second order volitions, and therefore was not responsible. On my account, he acted on a desire that he identifies with, and therefore he is responsible for his action. This seems to be more intuitive than the answer Frankfurt gives. One may argue in favor of Frankfurt by saying that at the time of adoption the doctor identified with the desire so he is responsible in a derived sense. But

unfortunately, Frankfurt cannot say that because he uses this example particularly to distinguish between second order desire and second order volition. If the doctor identified with the desire at the time of adoption, then there the difference between higher order attitudes and second order volitions gets diffused, which is something that Frankfurt wants to deny, since he focuses on the decisiveness of the commitment as what makes him responsible for an action, and not the way it is related to other intentional states. One may point out that this might be problematic because in this case, the doctor does not actually identify with the desire, because after all, the only reason he had that desire was for research purposes. So on my account, he would also not be responsible for taking the drug. I would respond to that by pointing out the fact that as soon as someone considers a desire to be his own, without considering it to be alien to him, we can consider it to be one that he identifies with. After all, it is only having the desire that would make him similar to an addict, and able to understand what it's like to be an addict. So maybe what we need for responsibility is not identification in the strong sense of the term, like considering that desire to be a defining feature of one's personhood for that leads into a circle. What is needed is a weaker sort of identification. This weaker identification only entails that one considers that desire to be his own, and forms an integral part of one's psychology, and not an essential part of it.[79] So as long as it plays the right sorts of causal roles,

---

[79] One may ask whether he can consider a desire his own but want to get rid of it. There are two answers to such a question. Either he genuinely does want to rid himself of it, in which case although he believes it to be his own, it is not really, but is actually alien to him. Or he does not really want to

and is related in the appropriate way to the rest of his distinctive psychology, then it is a desire that one identifies with. This is one reason why my account might help solve some worries that Frankfurt's doesn't.

One possible problem for this account is its epistemological fuzziness, namely that this account might not allow us to distinctly know when one is responsible for one's action. It seems unclear what knowing whether a desire is alien to someone entails since it involves something internal to the agent. But although this epistemological difficulty is one worth noting, it has no bearing on the metaphysical and normative worries that this solves, since what we are looking for are necessary and sufficient conditions for responsibility. Recall that our definition of responsibility involves social practice and so, just like many of our other practices, some epistemological fuzziness arises. So it might just be that the fuzziness is something forced on an account rather than a weakness in it. But we still need some account of what it means for one to consider a desire to be alien to him. All I can do for this purpose is present an analogy that I hope is helpful. When someone needs an organ to replace a defective one in his body, he might be granted an organ that happens to be incompatible with his immune system. In this case, his cells react by attacking this organ and trying to get rid of it. Similarly, when one's desire is incompatible with his character, he reacts to this desire in a way that brings up all the conflicting judgments and higher order attitudes that he has. His character, so to speak reacts by rejecting this desire as part

---

rid himself of it and is somehow deceiving himself. This has to do with how well it meshes with the rest of his distinctive psychology.

of it, and although this desire might persist, it never gets assimilated into one's character. This makes a desire alien. If the desire is not alien to a person, i.e. if he has a positive higher order attitude towards it, and he acts on it, then he is responsible for his action.[80]

Perhaps a word needs to be said concerning how this is supposed to work for one's character. One's character does not consist in a simple aggregation of higher order attitudes and other mental states, but it is more like a web of linked and interacting psychological states. Some desires of yours cause you to have higher order attitudes, and vice versa, while some higher order attitudes might cause you to have other higher order attitudes or other desires. Sometimes you recognize that two attitudes conflict and decide that it is reasonable to give one of them up. So one's character involves interacting higher and lower order states. For a desire to be alien, it has to be one that cannot get assimilated into this web of linked and interacting states. This makes the desire one that is not part of the person and gets rejected by his character. So if a desire that led to some action is one that is so alien to this web that it cannot get assimilated into it, then the action it led to is not one the agent is responsible for. The action is caused by something that is in direct conflict with his moral character. That is why it would not be considered reasonable for us to have a reactive attitude towards that person with respect to this action.

---

[80] One may ask whether one is responsible for knowingly acting on a desire that is alien. I believe that one is not responsible in this case. Recall that if a desire is alien then he really does not want to commit the act, and so even if he knowingly acts on it, he is not responsible for it.

One may point out here that this account has features that are similar to Frankfurt's, namely the higher order account, but is more demanding than his account. Frankfurt believes that one is responsible for his action if he has a second order volition in line with it. In other words, one has to be decisive about which desire is to be effective. I do not believe that we should be this generous in our excusing conditions. On my account, one is not excused unless the desire that led to the action was really alien to the agent. There must be a strict excuse for one to be exempted from responsibility. This is the driving motivation behind my account. The other driving motivation is mentioned above, namely distinguishing the conditions of personhood from the conditions of responsibility for particular actions. Now let us turn to some of the strongest objections to this account.

3. Some objections to Higher order Endorsements

a. Taking Responsibility vs. Being Responsible

One worry that might arise for this account concerns the distinction between being responsible for and taking responsibility for an action. Agents sometimes take responsibility for actions that they are not responsible for. At other times, agents are responsible for actions but do not take responsibility. Taking responsibility means considering oneself to be morally responsible for an action (even though they might not be), and therefore accepting the consequences of being responsible.[81]

---

[81] In his book *The Metaphysics of Free Will* (New York, Blackwell Publishing, 1996), J.M. Fischer gives an account of being responsible in terms of taking responsibility for his actions. He relies on the idea of guidance control and what that is like for the agent. One can make a case for that, but this is not

One problem that has been pointed out is that given my account, this distinction seems to collapse. It seems that on my story, one cannot be responsible for an action without taking responsibility for it, nor can one take responsibility for an action without being responsible for that action. This would mean that taking responsibility and being responsible for an action amount to the same thing. If this is a consequence of my account then I have a problem. Here a few examples may be helpful. I will first consider two examples in which an agent is not responsible but takes responsibility, one involving an action that someone else had done, and one involving an action that the agent himself had done. Then I will give an example of someone who is responsible but refuses to take responsibility.

Consider Bushra, a loving mother whose daughter committed a crime. Although Bushra is not responsible for the action, she may love her daughter so much that she confesses and takes responsibility for that action. In this case, it is clear that Bushra is not responsible but takes responsibility for the action, This is a case of one who has not even performed the action but takes responsibility for it. Another example is that of Adrian, who is waiting to meet his future in-laws by a newspaper stand. As he is waiting, he finds an issue of Playboy that he knows a friend of his had written an article in, so he picks it up in search for his friend's article, and his in-laws witness him looking at the magazine. Adrian decides that if he were to try to explain his action he will look like a liar as well as a pervert, so he decides to refrain from doing so and

---

the path I will take, because I cannot rely on guidance control for my account, and the next paragraphs will show why I believe we are fallible this way.

take responsibility for his action. In this case, Adrian was blamed for performing an action that he was not responsible for (looking at indecent pictures while waiting for his future in-laws) and he took responsibility for that action. The distinction in this case is pretty clear as well.

Now let us turn our attention to a case in which an agent is responsible but refuses to take responsibility for her action. For instance, Cyma is asked to baby sit her niece, and while doing so, she gets a phone call informing her of a show by her favorite band that is at a local bar. Knowing that a bar is an unsuitable place for an infant, she takes her niece and goes anyway, claiming that she could not resist the temptation and that no one can blame her for dragging an infant into a smoke filled room. Although we would hold Cyma responsible for her action, she may still refuse to take responsibility for the action, claiming that it was somehow out of her control, or that it was a lower level desire that she did not have any positive higher order attitude towards.

Let me consider the two cases. The first one is a case in which an agent takes responsibility although he is not responsible for his action. An agent might recognize that the conditions for responsibility do not apply to him, and realize that if he were to judge his own action then he would not find it reasonable to have some reactive attitude towards himself because he doesn't have a positive higher order attitude towards the desire that led to the action. However, he may find for some reason or another that he wants to accept responsibility for the consequences of this action. One can use the analogy to a game in which the player recognizes that according to the strict

rules of the game he did not lose a turn, but to save himself time and energy tells the other players that he accepts the consequences of losing a turn. For the purposes of responsibility, a new example may be helpful here. I will use an example that I had actually witnessed in my high school years. Walid was an upstanding and extremely hardworking young man. He was an ideal student and every teacher's idea of exemplary. One day, during recess, another student, more of a trouble maker, Karim, threw something at one of the teachers. The teacher was very upset, noticing that the students had been laughing at him, and went straight towards Walid, thinking him to be the culprit, and forcefully tried to drag him to the principal's office. Walid snapped and pushed him away screaming at him that he had absolutely no right to do that. This young man acted on a desire that is not at all part of his character, for he pushed a teacher who was his senior both in rank and age. However, Walid took responsibility for the action, and felt guilty about it claiming that he is responsible and will take whatever punishment he is given. Here is an example of someone who may have known that he was not responsible for an action, but decides to accept responsibility for the consequences anyway. The man who takes responsibility accepts that he is not responsible, and recognizes that it is not reasonable to have some reactive attitude towards his action but still accepts these attitudes.

The second case is where one is responsible for one's action but refuses to take responsibility for it. It is possible for an agent to understand when the conditions for being responsible usually apply but fail to realize that

they apply in his case. One may know when it is reasonable to have certain reactive attitudes in general without recognizing that they apply in one's own case. Here is a case that may involve some self-deception, like the one involving Cyma mentioned above. The analogy to playing a game also applies here. A player may know all the rules of the game but refuse to admit that a particular rule applies to him at a certain instance. People tend to be partial to themselves and fail to acknowledge that they have lost or have done something wrong. A person may even be in a situation in which the same thing happens to someone else and argue that in the case of the other person these rules do apply, but fail to recognize it in his own case. People are fallible and mistakes are to be expected of them.

The above discussion shows that on my account one can be responsible for an action without taking responsibility for it, and vice versa. These two are not the same and do not get confused on my account. We are interested in conditions that make one responsible for his actions, not the conditions under which he would take responsibility.

b. Why the Second Order?

Another objection to this account is that appeal to higher order attitudes is either circular or too trivial to be the basis of both personhood and responsibility. Although, as I mentioned above, higher order attitudes are to be distinguished from second order desires, the following argument may still hold against higher order attitudes. Dennis Loughery[82] distinguishes between

---

[82] Loughery, Dennis 'Second-order Desire Accounts of Autonomy' International Journal of

what he calls 'instrumental' and 'non-instrumental' second order desires. Instrumental second order desires arise when there is a conflict of first order desires, or when one wants to have a desire in order for a different desire of his to be satisfied. Loughery argues that these don't seem to be essential for one's character, and therefore cannot be what Frankfurt has in mind. If I wanted to impress my friends, and smoking did the trick, then I would want to desire to smoke. This would be an instrumental second order desire, but it would be rather strange to consider this second order desire to smoke as what makes up my character. He goes on to discuss non-instrumental second order desires arguing that he does not know what Frankfurt means by them. He gives a few alternative interpretations of what these second desires mean and dismisses all of them. His conclusion is that there is no useful account of autonomy and responsibility based on second-order desires. I will discuss the last alternative that he offers of non-instrumental second order desires and defend my account against his objection.

"The final way open to us – as far as I can see – is to evoke the idea of making judgments about first-order desires.[83]

In previous sections I defended the view that it is precisely these judgments that are essential for both personhood and responsibility. It is precisely these judgments that make one a particular person and that make him responsible for his action. These judgments make us understand what it is to be a person and why such beings are responsible for their actions. Loughery argues in

---

Philosophical Studies, vol, 6 (2), 211-229 1998

[83] Ibid p. 222

particular against Eleanor Stump's second-order account[84] which places emphasis on the intellectually approved desires. According to Stump"

"[T]he will is free with respect to a volition V just in case V is accepted by the agent because his intellect approves of V (at that time under some description) as the good to be pursued and there is no higher-order desire of the agent's with which V is discordant."[85]

This is a little different from my account because of the intellectual element of it, but Loughery criticizes it in a way that is similar to how he would criticize my account. He argues that since it is the intellectual approval that makes the act autonomous, there is no need to talk about second order desires. Once we talk about attitudes taken towards these desires, we lose the second order. In other words, what happens at the second order needs to involve a desire whose intentional content includes another desire, and not, Loughery believes, merely having some attitudes towards one's desire. According to him having an attitude towards one's desires does not involve anything of the second order, and we may be able to talk about it only by discussing desires of the first order. He also brings up the problem of infinite regress that second order accounts face, namely, the objection claiming that there must be a third order desire whose intentional content is the second order desire, and a fourth order desire whose intentional content is the third order desire, and so on. He claims that there is no reason why we should place much emphasis on

---

[84] Stump, Eleanore 'Sanctification, Hardening of the Heart and Frankfurt's Concept of Free Will' The Journal of Philosophy, Vol. 85, No. 8 (Aug 1988) p. 395-420

[85] Ibid p. 407

the second order and not the third or the fourth order, and this yields a problem involving infinite regress.

To respond to the first of his objections, it seems that there is a terminological disagreement. Loughery does not consider a case in which one takes an attitude towards some desire to be anything of the higher order, whereas I do. On my account, when we have an attitude about some first order desire, this involves recognizing that we have that desire, and having some intentional states whose intentional content includes the desire. So what I mean by an attitude's being of the higher-order is that it is one which the agent recognizes as a desire of his and adopts some attitude towards. If Loughery's point is merely a terminological one, as just mentioned, then that is not a problem for Stump or me. If his worry is that discussing the second order will inevitably get us to an infinite regress, and that Stump and Frankfurt's replies concerning the second order being outright or a decisive commitment are not convincing, then I'm somewhat sympathetic with this objection and so will give a different response. Relying on the decisiveness of these commitments needs a further argument, and cannot be the sole response to the regress problem. There has to be a reason that makes second order attitudes of some significance, and that is in need of some justification. On my account, an attitude is higher order just in case it has an intentional state as its object because it is an evaluation of this intentional state. This attitude towards the intentional state, whatever it is, grounds personhood and responsibility because it enables the agent to think of himself

as an agent. If an agent examines these second order attitudes and hence forms third order attitudes he is still engaged in the same type of activity, and this would also contribute to what makes him an agent. What is crucial for my account is that this higher order endorsement or attitude is of the higher order, no matter how high it gets. If something is a third order endorsement then it is just as interesting as a second order endorsement, and the same goes for the fourth order and so on. Infinite regress is not a problem for my account, but another feature of it. It is the fact that this higher order attitude is of the higher order that makes it interesting for our purposes. Since both Frankfurt and Stump both rely on the decisiveness of the commitment, they are forced into the problem of infinite regress. The causal dynamic of their views causes them to fall into this trap, since it is the decisiveness that makes one responsible for an action, and so they may have to have a decisive commitment for their decisive commitment, and so on. This is a problem that I do not face, because I do not need to go to another higher order to makes something more decisive. The richness of one's moral character has no end, and that deals with the questions of regress, and this helps escape the condition for going higher. One's moral character is made up of recognizing one's own intentional states and taking attitudes towards these states, and any attitude taken will be part of one's character. This is why the richness of the moral character solves the problem of infinite regress. That is also why I am able to avoid one of the biggest problems for second order accounts

without positing an arbitrary point of decisiveness that Frankfurt and Stump posit.

c. Coercion

An obvious problem that will arise for my theory concerns coercion. I will use Robert Nozick's[86] definition of coercion according to which an agent P is said to be coerced by agent Q into performing an action A if and only if P is justified in believing that Q would have made P significantly worse off by not performing A than P was before Q posed that threat, and this threat played a part in P's decision to perform A.[87]

In discussing coercion, let us take the following example: Someone points a gun to your head and threatens to kill you unless you perform some immoral action, like giving them your friend's money that happens to be in your possession at the time. One may argue against my theory by saying that what caused the action was a desire of yours, namely the desire not to be killed. The desire not to be killed is certainly one that you identify with, and therefore, according to my account, you are responsible for giving the robber your friend's money. This seems to be unacceptable to most people, and therefore my account fails. However, I will argue against this objection by relying partly on Frankfurt's[88] response to similar sorts of criticisms, namely

---

[86] Nozick, Robert, 'Coercion' from *Socratic Puzzles* p. 15-44 (Cambridge, Harvard University Press, 1997)

[87] I will take this definition and avoid discussing the controversy concerning what the definition of coercion is, or what the difference between threats and offers is. For more on this see Nozick and Frankfurt.

that you are in fact responsible in some sense for giving in to the coercion, but that you are excused because you have not done anything that is blameworthy. All you are responsible for is saving your own life, which happened to involve giving the robber your friend's money. Under one description, we can think of the action as giving away your friend's money, but under another description it involves saving your own life. You are certainly responsible for saving your own life, and no one denies that. But since it is the same action as giving the mugger your friend's money, then you should be considered responsible for that action as well.[89] We are justified in having some reactive attitude towards the victim of coercion. However, the reactive attitude that we are justified in having is not one of resentment, since you did not perform an immoral action. In other words, when you gave away your friend's money, you performed an action, and one which you are responsible for, but the action in itself was not something that we can reasonably feel resentment towards you for. If you gave your friend's money to a bank in order to pay off a loan without the consent of that friend, then that might be an action that a judge can reasonably resent you for. But in the case of being mugged, the action the agent performed was not an immoral action. He was

---

[88] Frankfurt, Harry 'Coercion and Moral Responsibility' from *The Importance of What We Care About* p. 26-46 (New York, Cambridge University Press, 1994). My response is only to the objection mentioned above, not the cases Frankfurt describes which yield in the agent irresistible desires to give in to the coercion.

[89] One may argue that you are only responsible for the descriptions of an action that you are aware of, and that you may be exempted from responsibility if you are unaware of a certain description. For instance if you open a can of soda that happens to contain a toxic chemical that kills everyone around you, your action of opening that can of soda can also be described as releasing the toxic chemical. It may be that you are responsible for opening the can of soda but not for releasing the chemical, since you were not aware of this description of the action. But this is a problematic issue concerning individuating actions, and I will not deal with it since it is outside the scope of this project.

responsible for what he did, but he is not blameworthy. If we were in his situation then we probably would have done the same thing.

One might ask: Wouldn't it just be easier if we include coercion under the excusing conditions that Strawson discusses? Why should we hold an agent responsible for this non-blameworthy action instead of saying that the action was itself wrong, but the conditions excuse him from being responsible for it? The answer to these questions concerns more radical cases of coercion. We might not want to blame the person who was mugged on the street for giving away his friend's money, but there are other cases of coercion that we might not be as lenient about. Consider the following case[90]: Someone comes to you and threatens to cut off your leg if you do not perform what they ask you to. They ask you to press a button that they themselves cannot press. The button that you are about to press will kill everyone in Africa. If being coerced was an excusing condition, then you should not be blamed for destroying Africa. Many of us would still blame you for giving in to the coercion. We expect people not to perform some actions, even if they were coerced.

People should not be exempted from responsibility only because they were ordered or coerced to do something. These cases are not always excused, because sometimes the actions agents perform under these conditions are too immoral to warrant excusing them. The best way of dealing with this problem, therefore, is to say that they are responsible in all these

---

[90] These cases are very similar to the ones presented by Frankfurt.

cases, but in some of these cases they are not blameworthy, and in other cases of giving in to the coercion they are.

Conclusion

In this chapter I have discussed my view of moral responsibility. I started the chapter by giving an account of the type of responsibility we are concerned with. I followed Strawson's view, and argued that one is responsible for one's action if he is the appropriate recipient of certain participant reactive attitudes. In order to have any participant reactive attitudes towards others, one has to be able to take the intentional stance. Adopting the intentional stance towards oneself enables one to understand what it means for one to be responsible for an action. We then discussed the conditions that make one responsible for one's actions. I argued that what matter are the higher order attitudes that one takes towards one's intentional states in general and his desires in particular. I also pointed out some problems with the most popular account of higher order views of responsibility, and in particular Frankfurt's hierarchical view. I argued that one is responsible for his action if and only if he has a higher order endorsement of this action, in other words if he recognizes the desire that led to the action and approves of this desire. Approving of a desire involves considering this desire to be one's own, and not alien to oneself. I responded to some problems with this account. In the next chapter, I will discuss the conditions that make one remain responsible for a given action over time, and see if these overlap with the conditions that make one the same person over time.

Chapter 4 Conditions of Personhood and Responsibility over Time

"This is a separate argument for the existence of an irreducible self. In order that we assign responsibility, there must be an entity capable of assuming, exercising and accepting responsibility. We will understand this point better if we introduce the notion of time. The notion of responsibility makes sense only if we can now assign responsibility for actions that occurred in the past. I am held responsible now for things I did in the distant past. But that only makes sense if there is some entity that is both the agent of the action in the past and me now. That entity is what I have been calling 'the self'. Notice that I am not in that way responsible for my perceptions. Perceptions affect me but I am not accountable for them in the way I am accountable for my actions."[91]

In this chapter I will discuss the different conditions for remaining the same person over time, as well as conditions of remaining responsible for an action over time. In the traditional literature, as the quote from Searle above suggests, writers have assumed that remaining the same person over time is a necessary and sufficient condition for remaining responsible for an action that the person has committed. I will argue against this equivalence, and I will try to give an account of what makes someone the same person over time and a different account for what makes someone remain responsible for an action over time. I will argue that while remaining the same person over time involves keeping a large number of mental states and higher order attitudes sufficiently similar, what makes a person responsible for an action is that he is

---

[91] Searle, John *Rationality in Action* (Cambridge, MIT Press, 2001) p. 89-90

psychologically continuous with the earlier person who had performed the action. The kind of psychological continuity that is required for this will be discussed in a later section of this chapter. So although it is obvious that the agent has to remain a person over time, he does not necessarily have to remain the same person as the one who performed the action for him to be responsible for that action.

1. Personal Identity Over Time

a. What Makes One a Particular Person at a Time

In Chapter 2, we discussed a distinction that Peter Unger[92] makes between core and distinctive psychology. One's core psychology, as we mentioned, consists in what a person shares with other people, whereas one's distinctive psychology consists in a set of particular psychological states that he shares with no one else. We said that the relevant part of the distinctive psychology for responsibility is the set of higher order attitudes that the person has. It is the fact that he approves of some of his desires and disapproves of others that makes him the subject of moral responsibility, and that makes him capable of holding other people responsible for their actions, and hence capable of understanding what moral responsibility entails. This is also what makes him part of the moral community, and these higher order attitudes, as I have argued in previous chapters, ground both personhood and moral responsibility. It is therein that one's moral character lies. What makes me a different person from you is that I have a different moral character, so

---

[92] Unger, Peter 'The Physical View' from D. Kolak and R. Martin (ed.) *Self and Identity* p. 192-213 (New York, Macmillan Publishing Company, 1991)

what makes someone a person (in our sense of the term) is that he has a character. One's moral character comprises of higher order attitudes which make one a particular person. They make up the part of our distinctive psychology that is relevant for responsibility. These higher order attitudes are what make one a particular being who can be held responsible for one's action at a given time. Since it is these higher order attitudes that ground personhood and responsibility, these are what make one a particular person at a given time.

I have argued for each of the conclusions of this section in Chapter 2. Although Unger makes the distinction between core and distinctive psychology, he does not believe that the same distinctive psychology is what is needed to make someone the same person over time.[93] I disagree with Unger, and that is what I will argue in the next section.

b. What Makes One the Same Person Over Time

Since what defines personhood is a set of higher order attitudes, what constitutes personhood over time must have something to do with that as well. One natural thing to conclude is that in order for one to remain the same person (in our sense of the term) over time, the later person has to have a sufficiently similar set of higher order attitudes to that of the earlier person. In other words, if the later person has enough attitudes that are different from those of the earlier person, then they are not the same person. If it is the distinctive set of higher order attitudes that makes one have a unique moral

---

[93] Ibid p. 195-199

character, then it must be this distinctive set that also makes him a particular person at a given time. If enough of these higher order attitudes change, then the character changes. If the character changes, then this person changes. Recall that what was distinctive about this person was his character, so if that character changes then we have a different person.

Now of course this sort of relation admits of degrees, for some person at time t1 might share more with a person at t2 than with a third person at t3 (assuming the paths connecting all three are continuous). We can say that if a later person shares at least half of the higher order attitudes with an earlier person, then he is identical to that earlier person. The more the later person share with an earlier person the more they are to be considered similar to that person. Consequently, three persons can all be identical to each other while there may be different degrees of similarity between them (so long as each retains at least half of the distinctive psychology of the others).

Obviously this is a strange sort of identity and we may even want to say that what we are doing here is abandoning talk of strict identity. Since this discussion of personhood is trying to answer a normative concern, we need not expect strict metaphysical rules to apply to it. Since I am going to argue that responsibility does not depend on personal identity, it does not matter whether we can give strict answers to questions of personal identity. We may not be able to find a relation to give us strict answers to questions about personal identity over time, but that doesn't mean that we cannot give strict

answers to questions of responsibility. This is where I believe that the confusion in this whole debate lies.

My view so far has been very similar to that of Parfit[94], but there are two major disagreements that I have with him. The first is that I place a lot of emphasis on the specific part of the person's distinctive psychology that is essential for a person persisting over time. I believe that one's personhood lies in his character, which is determined by his higher order attitudes. Parfit does not specify that any part of the distinctive psychology needs to persist for us to have the same person. This disagreement is probably due to the fact that I emphasize the normative concern, and my specific concern is what persons should continue to be held responsible for some action. Parfit is more concerned with survival, and whether I should care about future persons. Another major disagreement between my account and his will come up in the section on moral responsibility over time. Parfit believes that the relation between personal identity and remaining responsible for an action is much stronger than I do.[95] This is a disagreement that I will elaborate on in the section dedicated to it. But now let us look at some of what I consider to be major objections to my account of persons persisting over time.

c. Some Objections

Perry

---

[94] Parfit, D. *Reasons and Persons*. (New York, Oxford University Press, 1984)

[95] Ibid p. 323-326

As I mentioned in the first chapter, John Perry[96] criticizes views of personal identity over time that are based on identification and personality for a number of different reasons. Before getting into the actual objections, it is important to note that Perry's concern is mostly about whether I should care about a future person that is supposed to be identical to me. Although he believes that a later person will end up being responsible for the earlier person's actions (if the two are identical), responsibility is not his main concern. That is why while addressing his objections I will address concerns about whether we should be concerned with this future person. Now let us turn to these objections.

First, Perry argues that it is a contingent fact that persons change personality. Persons, on Perry's account, change personality too often for it to be the condition that grounds identity over time, and so change of personality cannot be the basis of changes in identity. Of course this objection on its own completely begs the question against my account. I have presented arguments for why I believe that changes in personality do result in changes of personhood. If it is persons in the forensic sense of the term that we are concerned with then character does shape the person, and changes in character do result in different persons, however frequently or infrequently this occurs.

Perry's second and stronger objection is the following. I will still have concern for a person that is psychologically continuous with me, even if she

---

[96] Perry, J. 'The Importance of Being Identical' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 67-90

has a personality that is drastically different from mine. If we apply this objection to my account, we would say that personal identity cannot be about higher order attitudes because you would still care about someone who is psychologically continuous with you even if that person has a different set of higher order attitudes from you. If someone were to tell you that in twenty years the person you would be psychologically continuous with will have very different political and moral views from you, and that person will have different friends and judge people differently, surely you will still be concerned with that future person's well being. Perry argues that this concern is to be explained by the fact that it is still you, despite all these changes. So too he might argue that this later person will be responsible for my actions because we are the same person despite the differences. I have two responses to Perry's objections.

First, most people do not radically change their character very often, and if you do change too often you are a very strange person. Although this happens to some people, it doesn't happen often enough for us to think it the norm. But of course this is an empirical question, and it cannot be my only response to such objections. We do tend to assume that people remain the same throughout their lives, and also remain responsible for their actions throughout their lives. Since we tend to assume these two (and since it is usually empirically true), then it seems reasonable for us to maintain that our criteria for personal identity and responsibility over time tend to coincide. But this is in fact not the case, which is precisely what I will be arguing.

Second, I would like to appeal to a thought experiment. Let us say that someone from the future visits you and tells you the following: starting tomorrow, a series of really unfortunate events will happen to you and will turn you into an absolute cynic. Actually, they don't just turn you into a cynic, but your whole character is completely compromised, and this resulting person becomes insensitive to suffering. At a later time, this person who is psychologically continuous with you (but has now become so drastically different from you) will actually start getting pleasure from the suffering of others and end up being a serial murderer, rapist and child torturer. This person becomes so good at getting away with his crimes that he gets a following and people start emulating his actions. This will inevitably happen, and there is absolutely no way that it can be avoided it unless you die. So this person from the future makes you an offer to kill you (or just confine you to some solitary place where you can never interact with the world). Again, the point is not whether you can choose otherwise or not, the point is that this is supposed to be you doing all those things, but would you think of that as you? Wouldn't you want to avoid being continuous with that person? Many of us will opt to make sure that we do not turn into this terrible person, probably because our characters are essential for us. Changing this character so radically threatens our very nature, and we would like to dissociate ourselves from this future person since we would not consider ourselves to be identical to him.

The reason I believe that Perry does not take changes in personality seriously is because he does not seriously take into account such radical changes in character. He thinks of other sorts of changes, such as changing your taste in music or choice of clothing. Even when people change their moral and political views, it is generally not because they come to value different things, but it is because there are many new facts that have been presented to them. It is not that often that people change a large number of their higher order attitudes. That is of course an empirical fact. But if it were the case that people do fundamentally change their attitudes towards their own desires, they would still dissociate themselves from a person who is continuous with them but is so radically different from them. They would not consider themselves to be identical to that person. In sum, I do not believe that our intuitions are on Perry's side.

Unger

Although it is Peter Unger[97] who introduced this distinction between core and distinctive psychology, he does not believe that distinctive psychology is what determines personal identity over time. Like Perry, Unger argues that one's distinctive psychology changes too often for it to be what makes up the same person over time. He concludes that what determines personal identity over time must be one's core psychology.

My response to this objection is going to follow the same line as my response to Perry. Since there are particular features of distinctive

---

[97] Unger, P. 'The Physical View' p. 193-194

psychology that make one the type of creature that is capable of being held responsible for an action, this must be what constitutes personal identity. This particular part of our distinctive psychology is what I have called moral character and since it is this moral character that makes one a person at a given time, it should be that this same character is what makes him the same person over time.

One question that Unger might engage is the following: What are the requirements concerning core psychology? Does one also have to keep the same core psychology for him to remain the same person over time? In my second chapter I have discussed the different elements that make up one's core psychology. It seems that having the same core psychology ought to be a requirement for having the same distinctive psychology. In the section on responsibility, I will argue that having the same core psychology is what is needed for someone to remain responsible for an action over time. My view is that a person is responsible for the actions of persons he is psychologically continuous with. Yet although I agree with Unger that having the same core psychology is necessary and sufficient for responsibility, it is necessary but not sufficient for personal identity, which also requires sameness of character.

Transitivity

One objection to Locke's account that also applies to mine is the one raised by Thomas Reid[98], namely that it leads to a contradiction because the relation that grounds personal identity is not transitive. Reid gives the

---

[98] Reid, T. 'Of Mr. Locke's Account of Our Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 113 –118

example of a man who can remember a relatively recent time in his life but not a more distant one. In other words, an old man at t3 can remember being a young man at t2 but cannot remember being a child at t1. However, the young man at t2 does remember being the child at t1. If memory is a necessary and sufficient condition for personal identity, the old man is identical to the young man. However, on this theory, since the old man does not remember being the child, then he is not identical to the child. But in this example the young man at t2 does remember being the child at t1, so the young man is identical to the child. But if the young man is identical to the old man, who is not identical with the child, then the young man is not identical to the child. This means that the young man is both identical and not identical to the child, so on Locke's account transitivity fails. One might apply a similar argument to my account of higher order attitudes. We can easily imagine Ahmad at three different times of his life, times t1, t2 and t3, when the he has different sets of higher order attitudes. At time t2, Ahmad has mostly similar higher order attitudes to Ahmad at time t1, but also to Ahmad at time t3, but Ahmad time t3 does not share much with Ahmad at time t1. So Ahmad at t1 is identical to Ahmad at t2, and Ahmad at t2 is identical to Ahmad at t3, but Ahmad at t1 is not identical to Ahmad at t3, and so this is a case of failure of transitivity as well. Since identity is supposed to be transitive, then this account might be considered insufficient.

Instead of giving up my account, I will give up the transitivity of the relation that grounds personal identity over time and think of this as a matter of

degrees. If the relation is one of this sort, then the problem above concerning transitivity does not arise, because the relation is not supposed to be transitive. So we can say that Ahmad at time t1 is the same as Ahmad at time t2 (since they share a large set of their higher order attitudes), and Ahmad at t2 is also the same as Ahmad at t3 (for the same reason), but Ahmad at t1 is not the same as Ahmad at t3 (since they do not share a large enough set of their higher order attitudes). In this case, transitivity fails, but that is not a problem for my account, since I do not claim this relation to be a transitive one. But although the relation that grounds personal identity is not transitive, I will argue that what grounds responsibility is.

This is not an original line. Derek Parfit[99] argues along those lines. Parfit argues that there is no reason why what we care about in persisting persons should be transitive. Strict identity is not what matters for survival, but it is a different relation. This relation is not transitive, and the rules that apply to strict identity do not apply to relation R, psychological continuity, which he claims is all that matters. Similarly, I claim that having the same (or a very similar) character does not need to be transitive either. So my view of personal identity is parallel to what Parfit believes matters in personal identity. So on my view, what makes one that type of being can vary throughout his life, and the relation between the different persons at the different times does not have to be that of strict identity. Although I am able to say that a later person is sometimes determinately non-identical to an earlier person, but still

---

[99] Parfit, D. *Reasons and Persons* p.262-273

responsible for their actions, so this is not a problem for me. Now let us turn to how this sheds light on what makes one remain responsible for an action over time. And this is where my major disagreement with Parfit lies. My view of personal identity is parallel to what he believes matters for personal identity, my view of persisting responsibility is different from that, and that is what I will turn to next.

2. Moral Responsibility over Time

a. What Makes One Remain Responsible Over Time

After discussing what makes someone the same person over time, I will now turn to what makes a person remain responsible for an action over time. Being the same person as the one who performed an action (and was responsible at the time they performed that action) is a sufficient but not a necessary condition for remaining responsible for that action. I will argue that as long as the being who performed the action and was responsible at the time continues to be ***a*** person, this person is still responsible for the action. In other words, if a person performed an action (and was responsible at the time they performed it), and that person was psychologically continuous with a different later person, then this later person remains responsible for that action. All that is needed for remaining responsible is to be connected to the person who performed the action in a certain way. This connection has to do with continuity of personhood but not sameness of personhood over time.

One major issue that needs to be discussed here concerns the sort of continuity or connection that needs to hold between the later person and the

earlier person. I will argue for a causal account of continuity, which I will call psychological continuity. So long as the higher order attitudes are causally continuous psychological states from the earlier person to the later one, then the two persons are continuous. This causal relation doesn't have to hold between the different higher order attitudes, but if the cause of a new higher order attitude is an older one, then that is sufficient (but not necessary) for continuity. The newer higher order attitude must be caused by some psychological state of the earlier person, and not necessarily a higher order one. This involves a direct psychological causal story with one psychological state producing another. If this happened in a natural way, and the later set of higher order attitudes was a causal product of the earlier person's psychology, then we can say that the later person is continuous with the earlier one. Two persons can thus be continuous even though they have a radically different set of higher order attitudes. So a person can be psychologically continuous to another without being the same person. Being psychologically continuous with the person who had (responsibly) performed an action is necessary and sufficient for being responsible for that action. This means that a person can be responsible for an action that a different earlier person with whom he was psychologically continuous had done.

The position I have just advanced is in need of an argument. Before presenting this argument, I would like to make a distinction between two ways in which one may talk about a higher order attitude.[100] The first way involves

---

[100] This might be considered analogous to a distinction presented by Marya Schechtman in

isolating it as one specific mental state, namely a higher order one, which is just a psychological state that happens to be about another mental state. We may dub this way of talking about them as talking about "isolated higher order attitudes". Another way of talking about this higher order attitude involves considerations concerning how it is related to other higher order attitudes and other lower order states. You may even say it is identifying where this higher order attitude lies within the web of one's intentional psychological states. It is important to note how this attitude affects and is affected by the other intentional psychological states and after all, it is the collection of higher order attitudes that make up one's character. We may call this other way of talking about them talking about "nested higher order attitudes".

For the question of responsibility, it is the nested higher order attitudes that are more interesting for us. Recall from the previous chapter that it is the relation between a desire and one's other higher order attitudes the makes one responsible for an action the desire leads to. If this desire is in great conflict with his other attitudes, then the agent is not responsible for his action. After all, this is what it means for a desire to be in conflict with his character. One can think of these higher order attitudes as forming a web and having certain effects on one another, as we mentioned in the previous chapter. Other first order psychological states are also part of this web, because they are affected by these higher order attitudes.

'Personhood and Personal Identity' The Journal of Philosophy, Vol. 87, No.2 (Feb. 1990) p. 71-92. Her distinction is about memories, and she argues that one cannot possibly have the same memories as someone else without at the same time having a lot of what it is to be that person. We may call this the difference between isolated and nested memories.

Since this way of conceiving higher order attitudes is relevant to holding one responsible for an action at a given time, a similar story should be told for remaining responsible for a given action over time. This set of higher order attitudes is not a mere aggregation of psychological states but rather a web of causally interacting psychological states. It therefore seems that the effects of intentional states on one another (in the direct causal sense) play a large role in forming the later higher order attitudes of the new person. Since the later attitudes that may make up the new character are so intertwined with, and causally connected to, the older ones, it seems natural for us to conclude that there is a strong connection between the later person and the earlier one. This earlier person, although he stopped existing, caused the later person to come into being, and it is his attitudes and desires (which led to his actions) that are the parents, so to speak, of the new person. So this later person comes to have the character that he has because of the earlier person's psychological history, and this later person would not have been the person that he is if it were not for the psychology of the person that he is continuous with. The later person's distinctive psychology is causally continuous with the earlier person's distinctive psychology, and the same core psychology remained throughout. Although the two are different persons, the later person is very closely tied to the earlier person. So this connection between the earlier person, that you were, and the newer person that you currently are, is a very strong one. Although this connection is not strong enough for

sameness of personhood, since the characters are very different, it is a good basis for continuing to hold someone responsible for his action.

At this point, the following question arises: Isn't it counterintuitive to hold a person responsible for something that another person had done? To this I will answer using this commonplace example[101]. Consider a case in which a man, Max, proposes to a woman, Diane, and showers her with diamonds and other fine things. The police later discover that all of the money which was used to buy Diane's jewelry was stolen, and that her fiancée, Max, was a thief. He is sent to prison, and she is asked to give some reparations to the people who had been robbed. It does not seem counterintuitive to say that Diane ought to give this back. Although she herself had done nothing wrong, it still seems fair for her to give reparation to those who were harmed by Max for her sake. She did not commit any crimes, but she may feel somewhat uneasy, for after all, her whole lifestyle and all she possessed she got from innocent people who had been robbed. Of course in this case none of us will think that she is responsible for stealing the money. She was unaware of Max's activities and would not have approved of them had she known. But it is perfectly reasonable for us to expect her to pay reparations for the victims of Max's crimes that she happened to benefit from. The causes of her extravagant lifestyle were petty crimes and robberies, so in some sense she did benefit from these misdeeds. People cannot deny that, and it is natural for Diane to feel shame and guilt for these acts that she did not commit. If one of Max's

---

[101] I would like to thank Bashshar Haydar for coming up with this example.

victims had become down and out due to that robbery, then one may argue that she has special responsibilities to alleviate their suffering, although she herself had never wronged them, only benefited from someone who had wronged them.[102]

Now we can turn to the trans-temporal case of a person who committed a crime, and changed drastically to become a different person. As mentioned in the argument above, this connection is so strong, much stronger than the case of the Diane and Max that it seems easy to say that we should hold a person responsible for another's actions since, after all, the relation between these two persons is a very close one. Also, the connection between the higher order attitudes of the later person and the earlier one is very tight, so now it is no longer like the case of Diane and Max but a case of past approval of certain desires. Since the earlier person is continuous with the later one, then there is a strong enough connection between the two to warrant holding the later one responsible for the earlier one's actions. It is in this psychological continuity that one remains responsible for another's actions. Now we can turn to the obvious question: why doesn't this also apply for remaining the same person over time?

b. Why Responsibility over Time Is Different From Personal Identity over Time

In the above section I have explained what makes one responsible for an action over time. I have also argued for the conditions that make one

[102] Haydar, Bashshar 'Special Responsibility and Cost', unpublished paper

remain the same person over time, which turned to be different from the ones for continued responsibility. One might still raise the following questions: Why is there no overlap? Why can't we just give up one of the arguments and say that the other presents us with the conditions for both responsibility and personal identity over time? I will try to deal with both these questions in this section. As with most work in this area, this response will consist in two methods. I will first present some thought experiments that support my position, and then present an argument to support the intuitions we have. It is important to note that I don't believe that the thought experiments are what provide us with the reason for holding this view, but they are useful in explaining the possibility of holding this view. Let us now turn to these thought experiments.

The first thought experiment involves an example from the news. Many people discussed the case of Stanley "Tookie" Williams, an old gang member. Tookie was charged with murdering four people and was sentenced to death. Later on in his life, while in prison, he was thought to have redeemed himself and spent his days writing pieces condemning gang violence. He was put to death in December 2005.[103] Many people argued that his life should be spared from the death sentence while others disagreed. However, in the debate, although many people claimed that he had become a different person from the young gang member two decades ago, no one claimed that he was no longer responsible for what he had done (if he had actually done it). Some

[103] For more on Tookie, you can see http://www.tookie.com/

of those who argued for leniency claimed that he never committed the crime in the first place. Others said that he should no longer be punished because he had redeemed himself and paid his dues. For the purposes of this work, it does not matter whether he was in fact guilty or not, or whether his life should have been spared or not. It seems that on my account, one can say that although he was a different person from the one who had committed the actions he was being executed for, he remained responsible for these actions because the later person is psychologically continuous with the person who had committed the crimes. We might even say that he was forgiven for his crimes because of this drastic change of attitude, and it seems intuitive to hold that position. Whether or not he still needed to be punished or rewarded would depend on the theory of reward or punishment that one holds, but that is not the subject of this work. More discussion of practical applications of this theory will come up in Chapter 5 which deals with cases of multiple personality disorder. Similar consequences will follow for these cases.

The second example I would like to discuss is brought up by Derek Parfit[104] when discussing personal identity and responsibility. He brings up an example of a Nobel Peace Prize winner who is somewhat elderly, and who admits that when he was twenty he had beaten a police officer while drunk. Parfit argues that this old man is surely not responsible for this action, since he had changed so much. There is barely a connection, on Parfit's account, between the Nobel Prize winner and the drunken young man. Parfit believes

---

[104] Parfit, D. *Reasons and Persons* p.326

that this is a good reason for us to deny responsibility in this case. This is a crucial point on which Parfit and I disagree. I do not believe that drastic changes in personality can exempt one from responsibility. We may be less inclined to hold the reactive attitude towards the old prize winner than the drunken young man the next day, but the reactive attitude would still be appropriate in the first case. One thing we might do to clarify our intuitions is to take a case that is a little bit more extreme and see whether our intuitions still apply. Think of a former Nazi official who becomes a very nice person and helpful to the community. This former Nazi official has changed a lot, and there are very few of his beliefs that he still holds, and he has never killed anyone since the end of the war. This doesn't exempt him from responsibility although he is a different person from the one who originally acted. We can think of the movie Music Box (1989) in which the daughter of a former Nazi official tries his case in an attempt to prove that he was not the infamous Nazi official that everyone was accusing him of being. He was an old man at the time of the trial, a loving father and a kind and gentle man, but at the end of the movie she discovers that he was in fact the same man. She is not as upset at him for lying to her as she is at the fact that he had done these terrible things he had been accused of. None of us would think that in this case she is being unreasonable and holding inappropriate reactive attitudes towards a person who had not done anything wrong. This case seems to be clearer than that of the Nobel Prize winner, but they are very similar in terms of their structure. So just as we are inclined to think that the former Nazi

official is still responsible, we should say that the Nobel Prize winner is responsible as well. We might, if we are very charitable, forgive him or excuse him from punishment. (A discussion of forgiveness will come up in the next section.) This, however, does not mean that he is not responsible for this action. This is a major disagreement between Parfit and me. Parfit ties responsibility to personal identity or degree of connectedness (which is what he believes matters) whereas I tie responsibility to psychological continuity alone.

Now it might be helpful to turn to the argument that is behind these intuitions. This argument can be understood better as an answer to the objection that was raised in the first chapter. It might be true that personhood as we define it involves higher order attitudes, and that these attitudes form character. But perhaps having similar higher order attitudes is not a necessary condition for remaining the same person. In other words, being the same person has nothing to do with having the same higher order attitudes, but has to do with psychological continuity. So perhaps both relations, that of personal identity over time and that of remaining responsible over time are determined by psychological continuity. In other words, why is there a separation between personal identity over time and remaining responsible over time?

In the discussions of personal identity in the literature, most theories can roughly[105] be grouped in one of two camps. One camp considers being the

same person is grounded either in similarities of content of later memories and earlier thoughts and experiences, or in similarities between earlier psychological states such as beliefs and memories, regardless of the causal continuity between these beings. For instance Parfit states:

"Relation R is what matters. R is psychological connectedness and/or psychological, with the right kind of cause.

I also claim

In an account of matters, the right kind of cause could be any cause."[106]

In other words, on Parfit's view, what matters for identity has to do with psychological connectedness regardless of the kind of causal story between the earlier person and the later person. It is about having the same psychology rather than about continuity, since this psychology is the most relevant thing for what matters in survival. On this kind of view, if there is a great degree of similarity between these two persons, then they are identical. On this account, it does not matter how the later person was formed (whether he is causally continuous with the earlier person or whether he was created by a tele-transporter). The kind of similarities that we should be looking for depend on the particular theory at hand. Let's call this group of theories "similarity theories".

---

[105] I say roughly since some may be in both camps, for instance David Lewis claims that personal identity is dependant on both similarity as well as causal continuity. See Lewis, D. 'Survival and Identity' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 17-31

[106] Parfit, Derek *Reasons and Person* p. 262

Another group of theories of identity is based on causal continuity and asserts that being the same person as another involves being causally continuous with that person. On this account, it does not matter whether the later person is similar to the earlier person or not; what matters is that the earlier person is causally continuous with the later person.[107] Let us call this second group the "continuity theories".[108]

Let us turn back to my account. On this account, personal identity over time involves relations of the first and second sort, whereas remaining responsible for an act over time involves only the second sort. Now the objection above can be raised as follows: since the second sort of relation gives us continued responsibility, then it might consequently give us personal identity over time. The critic might demand a parsimonious theory. Either you have to be a continuity theorist for both responsibility and identity, or a similarity theorist for both. Showing that remaining responsible for an action involves causal continuity might also show that remaining the same person over time involves continuity. So if one wants to stay consistent then he will have to say that we have the same responsible agent. If this objection holds, then there is one thing that remains the same over time, and that thing is what remains responsible over time. The separation between personal identity over time and remaining responsible over time collapses.

---

[107] Examples of such theories are presented by Bernard Williams in 'The Self and the Future' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 179-198, Peter Unger in 'The Physical View' from D. Kolak and R. Martin (ed.) *Self and Identity* p. 192-213 (New York, Macmillan Publishing Company, 1991), and Eric Olson in 'Was Jekyll Hyde?' *Philosophy and Phenomenological Research* Vol. LXVI, No. 2, March 2003 p. 328-348

However, I do not believe that this case is a strong one. Recall that our major concern when discussing personhood and personal identity over time is responsibility and the type of beings that are responsible for their actions. We are ultimately interested in what makes persons remain responsible for their actions. The discussion of persons and personal identity over time is only instrumental for answering our real concern, which is responsibility.

My claim is that the conditions that make the same person over time are not the same as the conditions that make one remain responsible for an action over time. In some cases, we may find out that a person is responsible for an action that a different person had done. In the literature on personal identity, writers have assumed that the conditions of personal identity over time must be the same as the conditions of responsibility over time. I believe that this is false, and that is precisely what I shall argue for. Once we have understood what a person is, and once we have spelled out the conditions of what makes one the same person over time, we realize that they are not the same as the conditions for being responsible for a given act over time. If we abandon this assumption, we realize that many of the problems that arise concerning personal identity are dissolved, and many of the problematic cases become easy to discuss. This is what I hope to have shown.

Someone may say that my arguments demonstrate that persons (in the way I define them) need not persist for them to remain responsible for their actions, but something else persists, perhaps something like a responsible agent. In other words, the conditions for having the same responsible agent

and those of remaining responsible for an action are the same. So the same thing (this responsible agent) is precisely what remains responsible for its actions.[109] This weakens my thesis because identity over time is still considered a crucial element for remaining responsible for an action. But this completely begs the question against my suggestion. You are defining a responsible agent to be whatever remains responsible for an action, and that has nothing to do with personal identity anymore. You are saying that two things are the same responsible agent if and only if one is responsible for another's action. This does not help us answer any questions concerning responsibility, but actually relies on an account of responsibility. The question we are interested in for its own sake concerns what remains responsible for an action. Identity claims should be thought of only instrumentally.

What I have tried to establish is that we do not need to have the same person for responsibility. Sameness of personhood does not tell us anything about responsibility, and if you separate the two questions then things become clearer. It is sometimes true that questions of personal identity are vague and indeterminate and this is especially true when identity is based on similarity relations including my own theory. There will be some cases where there isn't enough similarity to ground personal identity, and sometimes it is genuinely an open question. That is one good reason why we should not rely on having the same person to have continued responsibility over time. We are also far more willing to accept that someone has changed to become a

---

[109] I would, again, like to thank Ellen Fridland for pointing this to my attention.

different person than to exempt this later person from responsibility. I believe that this is justified, because what we mean by person is defined in terms of higher order attitudes and moral assessments, and if these change, then what is distinctive about the person changes, and therefore he becomes a different person. However, responsibility cannot be so easily abandoned, since our reasonable reactive attitudes will remain reasonable despite the fact that the later person has different higher order attitudes from the earlier person, since this later person is psychologically continuous with the earlier person. That makes keeping our reactive attitudes towards the later person completely reasonable.

One of the virtues of my account is that we don't need to answer questions of personal identity to determine responsibility. But although this is the case, it does not mean that questions about personal identity are not important, but they are important for different reasons. We make reasonable assumptions about people's sets of higher order attitudes, and this helps us understand and predict their behaviors. It determines the individual's dispositions and character. So questions of personal identity have useful roles other than for determining responsibility, and that is why separating the two questions is crucial. Also, it is clear that we need to have some understanding of what a person is, and how they persist over time for us to understand their moral development and changes in their characters. But changes in character do not seem, intuitively or otherwise, to determine moral responsibility, since the latter is not a matter or moral psychology, and is not used for explanations

of one's behavior but determines reactions to past actions. The epistemological difficulties that arise in determinations of personal identity, therefore, do not arise when determining responsibility. This is why I believe my account solves many of the problems that other accounts of personal identity fail to, since these accounts fail to make crucial distinctions concerning the reasons we ask certain questions. We need an account of continued responsibility in order to be able to give praise or blame to people, but we need an account of personal identity in order to understand one's character and that may help us understand and predict their behavior better. Thus the two questions are separate, with separate interests in mind.

c. Forgiveness

Given these rather strict conditions on what make someone responsible for a given action over time, one might wonder what we should say about forgiveness.[110] There are two questions that need to be answered: first, what is one doing when he asks for forgiveness, and second what does one do when one forgives another. I'll answer both of these questions.

When one asks for forgiveness, he is first of all recognizing that what he has done is wrong. He is accepting that in this case he is responsible for his action, and he is conceding its wrongness. One can even say that he takes the right sort of attitude with respect to the desire that led to that action, while he did not have this attitude at the time he did it. In other words, when a person asks for forgiveness, he acknowledges his responsibility, and he

---

[110] I would like to thank Kevin Murtagh for asking me about this.

recognizes that he is the appropriate recipient of some reactive attitude. Second, and after this recognition and adjustment of attitude, the person now asks you to suspend that appropriate attitude towards him, maybe in return for something else, like a punishment you impose on him, an admission of guilt, or compensation. What is offered in return of course depends on the severity of the action, the reactive attitude attached and the relationship between the culprit and the person he is asking for this forgiveness. In a nutshell, when a person asks you for forgiveness, he is admitting guilt of the crime, and asks you to suspend the reactive attitude that is appropriate to hold towards him.

At this point it becomes easier to answer the second question. Forgiving someone means that although you recognize that they are still the appropriate recipients of some reactive attitude, you suspend that reactive attitude. So you know that it is reasonable to feel resentment towards them, but you suspend that feeling of resentment anyway. Although sufficient reparation has not been paid, and sufficient punishment not administered, you suspend your reactive attitude. This does not exempt someone from responsibility, it just changes the feelings you have towards him.

3. Memory: Why It's Not Enough for Either

Locke believed that what makes something a person is that he has the ability to be conscious of himself as himself. To be the same person over time, one must have access to past mental states or memories of himself as himself. Because Locke was only interested in this question for the purposes

of responsibility, he considered having memory of a past action (in other words having access to the mental states that a person had while he was doing the action) as a necessary and sufficient condition for remaining responsible for that act. My account is somewhat different from that, because what is required of one to remain responsible for an action is not that he has access to past mental states, but that he is psychologically continuous with the person who originally performed the action. It is not clear whether or not this requires having memory of the past actions. It seems on the one hand that this type of psychological continuity might require some recollection of past actions, but on the other hand one can have an account that makes no reference to that at all. As we mentioned above, psychological continuity is defined in terms of a causal relation between these various higher order attitudes and other mental states. It may be that the memory of the past action is only relevant insofar as it plays a role in this causal story. In other words, if the memory of a past action had something to do with the causal chain that led to these higher order attitudes, then we can say that memory is relevant for remaining responsible for an action. In other words, memories are essential not for their content, but for the causal role they play in forming higher order attitudes. Although we are tempted to exempt people from responsibility if they have complete amnesia, it is not because they don't have memory, but is because there is no longer any psychological continuity. Memory is relevant only due to the causal role it plays in forming what is crucial for us, namely the higher order attitudes. The content of these

memories has no bearing on moral responsibility, nor does it determine personal identity. A more detailed explanation of this will follow in chapter 6, where I will discuss some cases of memory and explain the intuitions we have in cases of memory loss and their relevance for both personal identity and responsibility.

It is crucial here that if we are to count on memory as one of the conditions that make one remain responsible for a past action, it is only because of its causal role in forming the later person's higher order attitudes. So there is nothing in memory itself that makes one the same person over time, nor is there any direct connection between remembering an action and remaining responsible for it. Our general ability to remember is crucial for forming our higher order attitudes, but the recollection of some particular action or event does not determine whether you are responsible for that action or not. Perhaps some examples here might shed light on this. Consider a case in which you make a comment that hurts someone. You intentionally made that comment, knowing that person might be offended at it. A few days after this event takes places, he tells you about it. You do not recall making that comment. After all, to you it was just another comment that you had made, but you are still the same person who did it. The fact that you don't remember it does not make you a different person from the one who did it. Even if that person reminds you of this incident years later, and you are a different person, you still apologize and take responsibility for it. We never think that someone is not responsible for something simply because they do not

remember doing it. Also, if there is anything that psychology has taught us about repression it is the fact that we tend to repress memories that we do not want to have. This has nothing to do with whether we are responsible for them or not. Killing someone might be so traumatizing that conceivably one may not even remember doing it. But that does not exempt one from responsibility for the act. A longer discussion of this will follow in Chapter 6. We can also conclude that the memory of an action does not tell us anything about personal identity, because it does not determine an agent's moral character. Looking at the cases of multiple personality disorder might shed some light on this issue. That is the subject of the next chapter.

Conclusion

Given our conclusions from the previous chapters concerning what makes one a person at a given time and what makes him responsible for an action at a given time, I have argued in this chapter that what makes one remain the same person over time involves keeping a similar character. Remaining responsible for a given act over time involves being psychologically continuous with the person who had originally done it, regardless of whether the two are the same person or not. It is often assumed that to remain responsible for a given action, you must be the same person as the one who actually did it. I have argued against this assumption in this chapter. In the next chapter I will give a practical example of how this account can be useful for us. We will see the implications of this view on cases of multiple personality disorder. Let us turn to that.

Chapter 5 Multiple Personality Disorder: Some Cases and Their Implications

In the preceding chapters I gave an account of personhood and responsibility over time. We have realized that the same human being can develop to become a different person over time, yet the later person can still be held responsible for the actions of the earlier person. We have tested our intuitions with respect to some trans-temporal cases. I will now turn attention to cases of possible synchronic multiple personhood. In this chapter, I will discuss cases of multiple personality disorder in psychology, relying primarily on the famous case study by Morton Prince[111] of Christine Beauchamp and other persons incarnated in the same body. A discussion of this issue should give us a practical test of our intuitions, and hopefully clarify some points for my account.

1. Defining Multiple Personality Disorder

It is sometimes unclear what we mean when we talk about multiple personality disorder. Since we are going to be discussing cases of this disorder, we should have some working definition. For the sake of the current discussion, I will rely on what psychologists and psychiatrists today consider to be criteria for diagnosing a case. In other words, the following are considered conditions that may help us diagnose such cases, but are not necessary and sufficient conditions for the presence of the disease.

---

[111] Prince, Morton *The Dissociation of a Personality* (New York, Greenwood Press, 1969 Reprint)

According to the Diagnostic and Statistical Manual IV in 1994 of the American Psychiatric Association, one should look for the following:

"A. The presence of two or more distinct identities or personalities or personality states (each with its own relatively enduring pattern of perceiving, relating to and thinking about the environment and the self).

B. At least two of these identities or personality states recurrently take control of the person's behavior.

C. Inability to recall important personal information that is too extensive to be explained by ordinary forgetfulness.

D. The disturbance is not due to the direct physiological effects of a substance (e.g., blackouts or chaotic behavior during Alcoholic Intoxication) or a general medical condition (e.g., complex partial seizures). Note: In children the symptoms are not attributable to imaginary playmates or other fantasy play."[112]

These criteria, however, are not conclusive. They only serve as a helpful guide for diagnosis. For instance, there have been reported cases in which one of the personalities recalls what the other personality has done, like the case of Sally and Miss Beauchamp that we will discuss at length in this chapter. The manual takes note of that fact, and recognizes that there might be some cases that are anomalies. However, for our purposes, this will serve as a general guideline for recognizing such cases, rather than providing conclusive conditions.

---

[112] Diagnostic and Statistical Manual for Mental Disorders IV-TR Fourth Edition (New York, American Psychiatric Association, 2000) p. 487

There are many skeptics who believe that these criteria allow too much. These skeptics argue that some patients who do not have this disorder are diagnosed with it because the conditions are too inclusive. For the purposes of this work, I will not deal with the ongoing debate of how to diagnose people with this disorder, and depend on the criteria that are offered to us by the manual, as well as information given from some reported cases. I will also briefly discuss some claims made in the DSM IV concerning the possible reasons for the emergence of this disorder. But the issue we are interested in is whether this disorder may be able to generate different persons in the same body. Is it possible, in other words, to have more than one person, in our sense of the term, in the same body? This is what we will be discussing next.

2. Sally and Christine Beauchamp

According to my account of personhood it is possible for the same human being to change from one person to another over time. What is required for this change is that he develops sufficiently different higher order attitudes from the ones he previously had. Now one interesting question that might arise concerns cases of patients with multiple personality disorder: Is it possible for the same human being to be two different persons at the same time? According my theory of personhood, we may give a positive answer. In fact, the case of Christine Beauchamp that I will be discussing will provide many reasons in support of this answer. If we were to rely on my account of

personhood, we would have to say that this human being is several different persons.

Christine Beauchamp was a hardworking and reserved person who suffered from bad health.[113] At some point in her life, another person by the name of Sally emerged in the same body, and Christine was unaware of that. Sally had a completely different character from Christine and did not approve of many of Christine's beliefs and actions. Sally was careless and was more interested in amusing herself than in working hard and being a good Christian as Christine was.[114] Sally behaved like a child rather than a young lady, and had childish interests and concerns. Many times there were struggles between the two characters and they were both appalled by one another's characters and actions, and each of these personalities did not wish to identify herself at all with the other. In fact, Sally did some things that she knew would irritate Christine and even tried to harm her sometimes.[115] For instance, Sally picked up smoking, a habit she knew that Christine detested and so after Christine would take control of the body again, she would find the taste of cigarettes in her mouth and would get very upset.[116] Sally also made appointments with people that Christine would not have wanted to see, in particular with a man that Prince called "Jones" whom Christine did not want to associate with at all. Sally would make appointments with him, and then

---

[113] Prince (1969) p. 9-10

[114] Ibid p. 144 - 156

[115] Ibid p. 119 - 144

[116] Ibid p. 55

make sure she was capable of carrying them through.[117] Sally rebelled against her therapist, Dr. Prince, while Christine desperately wanted his help.[118] Sally also played a lot of practical jokes on Christine, who was much more innocent, for instance Sally once hid Christine's money and gave her a limited weekly allowance.[119] Christine never did anything to harm Sally, but was only upset at the consequences of Sally's pranks. These struggles went on for quite a while.

After the struggle had been going on for some time, a third character emerged whom Prince called BIV.[120] The reason he called her BIV was the following: The person whom he had been calling Christine was now called BI, while the hypnotic state of BI was called BII, and BIII was reserved for Sally's hypnotic state. At first Prince thought that this new character, BIV, was the same person as BII, the hypnotic state of BI. Later on, Prince realized that the new character that emerged was not a hypnotic state and had a different personality from Christine. Although there were some similarities between this new character and the reserved BI, there were many differences as well. BIV was not a child, nor was she only concerned with amusing herself, but she had sophisticated tastes and interests. She was an adult, like BI, but she had a different personality from BI, for instance, she was much less patient and

---

[117] Ibid p. 110-111

[118] Ibid p. 96-99

[119] Ibid p. 156

[120] Ibid p. 171-180

tolerant than BI.[121] She ended up fighting with Sally more than BI did, because BI was not confrontational at all, while both Sally and BIV did things that were harmful to one another. BIV wanted to get rid of Sally and did not approve of her actions and the difference between her and BI was that she was active about it. Sally also did not approve of BIV's actions, but thought it was a game. When Sally got access to BIV's mental states and realized that BIV really wanted to harm her and get rid of her, she was very upset. In fact, she became scared of BIV and considered her to be evil and she asked Dr. Prince to save her from BIV.[122] Prince describes the various differences between BI and BIV in a long list that includes differences in minor things like tastes as well as major disagreements concerning some moral issues. There were certainly some desires that BIV identified with that BI did not.[123] Prince compiles an exhaustive list of such cases,[124] for example BI *"enjoys doing charitable or altruistic work, visiting and reading to invalids or old people, visiting the poor, etc."*[125] while BIV "*Hate[s] such things"*[126]*.* There were many disagreements like this on the long list, and not just in matters of tastes but in assessments of actions.

---

[121] Ibid p. 197-200

[122] Ibid p. 312

[123] A problem might arise concerning how to identify these desires, and whose desire each one is, given that these characters are all instantiated in the same body, but I will say more about that in a later section. For now, let us focus on the differences in characters and look at examples of desires that BI identified with while BIV didn't or vice versa.

[124] Ibid p. 288-294

[125] Ibid p. 293

[126] Ibid p. 293

All three characters were clearly members of the moral community. They were all able to identify their desires and endorse them. They were also able to identify other people's desires and take attitudes towards them. They understood what it entails for each to be responsible for her actions, and what it means to hold others responsible for their actions. In fact, one of the struggles between these characters occurred when BI blamed Sally for losing her ring and Sally attempted to explain herself.[127] Another case in which Sally took responsibility occurred when Dr. Prince told her that he had to etherize her because of the bad thing she had done, and she begged him not to do it and asked him to explain to her what she had done wrong. The reason he was going to get rid of her had nothing to do with any particular thing she had done, but was only that he believed that she was not part of the formula for getting the original Miss Beauchamp back. He argued that the character of Sally was the result of the emotional distress and confusion of the original patient, Miss Beauchamp. He believed that Sally was not a part of the original Miss Beauchamp who had existed before the traumatic experience that led to the multiple personality disorder. The two other characters, however, showed signs of being parts of the original person, because of their preferences and their memories. This is why he believed that it was the synthesis of BI and BIV that would get him the original Miss Beauchamp.[128] Another interesting case in which one of the characters takes responsibility is described at the

---

[127] Ibid p. 189

[128] Ibid p. 398-405

end of Prince's study. After the original Miss Beauchamp returns, she takes responsibility for the actions of both BI and BIV, because she believes that: "after all, it is always myself".[129] Of course I am not committed to her claim that she is identical to both BI and BIV, but I do agree with the intuition behind her claim, namely the one concerning the fact that she is responsible for the actions of both characters. Since she is continuous with both of them, and it was the fusion of these two characters that resulted in her appearance, she is responsible for the actions of both of them. So although she is not the same person as either of the two, she is responsible for the actions of both. Note that this is different from her relation to Sally, with whom she is not continuous, since Sally was etherized and was not part of the fusion of characters. That is how she is different from BI and BIV. So although the later Christine Beauchamp is responsible for the actions of BI and BIV, she is not responsible for those of Sally. In a later section, we will look at how to think of these things over time, and the relations between the different characters over time.

Another noteworthy point is that Dr. Prince himself talks about Sally as though she is a distinct person from BI and BIV. He even says that he feels that neutralizing her seems to be "cold-blooded", since he had developed a relationship with her.[130] It was as though he was killing a person to make the original Miss Beauchamp return, but he believed that it was a worthy cause.

---

[129] Ibid p. 525

[130] Ibid p. 398

Whether or not that is true is not our interest, since we are concerned with whether these are distinct persons or not. Not only do these persons have the capacity to treat others as persons, but they can be treated as persons themselves which means that it is possible to take the intentional stance towards them. Each one of these characters is a distinct intentional system with a unique set of higher order attitudes and so it seems perfectly reasonable then to consider each of them to be a distinct person. This is a very commonsensical view, since Prince, and anyone who interacts with these distinct characters treats them differently and also thinks of them as distinct persons.

But we must note here that this case is different from trans-temporal cases, since in the trans-temporal cases there is psychological continuity between the earlier person and the later person. So if, for instance, Khaled was a different person in his earlier life than later on in his life, the two persons, the earlier Khaled and the later Khaled will be psychologically continuous with one another. This is not the case for BI and Sally, since they have distinctive psychologies that are not continuous with one another the same way the later Khaled's psychology is continuous with the earlier Khaled's distinctive psychology. This, of course, is a very crucial difference between the two cases.

One interesting fact about Christine and Sally is that Sally had conscious awareness of Christine's thoughts and actions although Christine was completely unaware of the existence of Sally. Even later on, when

Christine became aware of Sally's existence, it was through inference, unlike Sally's awareness of Christine, which was direct. Sally had access to Christine's mental states, would report to Dr. Prince what Christine was thinking, and sometimes even remembered what happened with Christine better than Christine did herself.[131] This creates a serious problem for the memory account of personhood and responsibility. If someone were to maintain that memory was a necessary and sufficient condition for remaining responsible for an action, then they would be committed to the strange claim that Sally is responsible for Christine's actions (because Sally remembers these actions), whereas Christine is not responsible for Sally's actions (because Christine does not remember these actions). This also poses a serious problem for Locke's account in particular, because Locke specifies that it is not just memories, but having access to the past consciousness of the person that makes a later person the same as the earlier one. In this cases, Sally is the same person as Christine (since Sally has access to Christine's past mental states), whereas Christine is not the same person as Sally (because Christine does not have access to Sally's past mental states). My account solves that problem, because Sally is responsible for Christine's actions only if her desires caused the action and she identified with these desires, or if she was psychologically continuous with Christine. Neither of these is true, first, because Sally does not endorse many of Christine's desires, and second, because Sally is not psychologically continuous with

---

[131] Ibid p. 31-33

Christine, since it is not the case that Sally's later higher order attitudes were causally continuous with Christine's earlier psychology. Because these two are different persons, and their distinctive psychologies are not continuous with one another, I can easily say on my account that Sally is not responsible for Christine's actions, and Christine is not responsible for Sally's actions. It does not matter whether one remembers the actions of the other or not, because memory, on my account, is important only insofar as it plays a role in forming the later person's distinctive psychology, and not an essential feature in its own right. A critic of my view would have to argue for why Sally's current higher order attitudes were directly caused by Christine's psychology, but this is very doubtful.

a. Identifying Mental States

One problem that arises for this story concerns how we can identify desires and recognize to which character each of these desires belongs. This may be difficult given the fact that these different persons are instantiated in the same body. Any materialist will have a problem trying to distinguish between the different mental states for the different characters. In this section I hope to give a satisfying solution to this puzzle. For the sake of simplicity, I will only talk about how to distinguish Sally's mental state from BI's, leaving those of BIV aside. Similar conclusions will of course apply to BIV's mental states.

Although Sally and BI have different personalities, there are some lower level states that these two persons share, such as the desire to eat or

drink. The same desire can be said to be both Sally's and BI's, and an action resulting from this desire is one of both agents. These two characters may share some of their distinctive psychology, and have some desires in common. The fact that they have some desires in common, or that the same desire can be considered as one that belongs to both of them does not seem problematic. But what should one say about cases in which both persons have a higher order endorsement of some common desire? I have to say that they are both responsible for this action, whatever that action is. If the action is amoral, like eating, we can very easily hold them both responsible for it. Whereas if a case is immoral, like murdering someone, and the different persons shared that desire which they both had a higher order endorsement of, then I think we have a good reason for holding these different persons responsible for the action. This would be as though they were both accomplices to the crime, but they are even closer than accomplices, given the fact that they actually share that desire that led to the action. Both Sally and BI would have been responsible for killing Dr. Prince if they both had a higher order endorsement of the common desire to kill him. However, if one endorsed the desire while the other did not, then the endorser would be responsible but the other not. It does not matter who was "in charge", since what it means for one of the characters to be "in charge" is for her desires to be effective. This squares well with "The Real Miss Beauchamp's" intuition that she is responsible for both BI and BIV's actions.

Now a more interesting problem arises: since some of the desires are not common for both persons, and these persons are incarnated in the same body, what makes a desire solely Sally's or solely BI's? Surely it cannot be that they identify with it, because that would mean that each gets to have only desires that she has a higher level endorsement of. There must be a different way of determining which desires are whose, otherwise I would be committed to the peculiar consequence that each of these characters is always responsible for anything she does, and that they are special in that they can pick and choose which desires are theirs. So what I need to say about this has to be something different. A desire is considered to be solely Sally's if it is connected to the rest of Sally's distinctive psychology in the right way. This desire, or thought, has to be causally connected to the rest of Sally's distinctive psychology, and is able to affect other thoughts or desires that Sally has. So any thought or desire of Sally's is part of that person's distinctive psychology, and not just the distinctive psychology that includes higher order endorsements. It has to be appropriately connected to the rest of Sally's mental states. This is somewhat similar to the definition given in Chapter 3 of what it entails for one to have a higher order endorsement of an action. But the key difference here is that this connection is weaker, and it does not have to be one that the agent acknowledges. Sally does not have to recognize that this desire is causally connected to the rest of her distinctive psychology for it to be hers. This desire might affect other lower level desires without Sally recognizing it or approving of it at all. As a matter of fact, she

doesn't even have to recognize that the existence of this desire at all. For her to have a higher order endorsement of the desire, she must recognize that she has this desire, and must have either a positive or negative attitude towards this desire, and perhaps try to encourage or resist the influence of this desire.

But now the following question might arise: Is it possible for a mental state to be distinctively Sally's without her having a higher order endorsement of it? It seems that we had defined having a higher order endorsement of a desire in terms of its connection to other states. Sally may have a desire that she does not have a higher level endorsement of if that desire is an unconscious one. This desire may be one that effects her other mental states, but one that Sally herself is completely unaware of. It is possible, therefore, for Sally to have a desire and act on that desire without being responsible for her action, namely if that desire is one whose presence she remains unaware of. It is impossible for one to have a higher level endorsement of a desire unless one recognizes the existence of this desire. However, that is not to say that if one recognizes the existence of the desire and one has a higher order endorsement of it then one recognizes the fact that one has a higher order endorsement of it. Thus one may deny that one has a higher order endorsement of a desire when one does in fact have one. But it is impossible for one to have a higher order endorsement of a desire when one doesn't believe that one has that desire at all.

In the discussion of this case, we have talked about the conflicts and struggles between Sally, BI, and BIV's different higher order attitudes. Their disagreements do not only concern trivial matters, but concern major issues as well. There are many cases in which each completely disapproves of the other's desires. My account can deal with that by claiming that although all characters are incarnated in the same human being, there are three distinct sets of higher order attitudes and three different persons there. Since what would make any two persons different is the fact that they have different distinctive psychologies, we can conclude that the different characters incarnated in the same body are also different persons. Although my answer involves appealing to my theory only, it may require some further discussion. We need to look at arguments in support of the fact that there is synchronic multiple personhood in these cases. I will rely on some arguments by Kathleen Wilkes[132] and defend these arguments.

b. One body, one person

Before giving reasons for the possibility of synchronic multiple personhood, Wilkes discusses another view. One may argue that since there is only one body that holds these different characters, then there must be one agent or one person. Since there is one action resulting from this body, then there must be one unified agent in this body, and that unified agent is one person. Whatever occurs inside the body in terms of reflection concerning the eventual course of action, the result is one unique action, and that must mean

---

[132] Wilkes, K. 'Fugues, Hypnosis and Multiple Personality' from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company,1991) p.115-134

that there is one unified agent. Wilkes believes that this is not a good argument because appealing to bodily criteria is not enough of a reason to consider all these different characters to be the same person. However, defenders of the bodily criterion can make a stronger case. They can say that differences in character or personality do not provide us with good reasons to believe in synchronic multiple personhood because there are cases in which similar conditions apply that we would not want to count as having more than one person in the same body. Consider social roles. Filling different social roles, each of us acts differently in similar situations. An explanation of this might be helpful.

For instance, when I am a modest student at the Graduate Center who is trying to learn as much as possible from people I play a different role from when I am a teacher to whom students run for advice and expect to teach them. When I go back to Beirut I play a different role from that of the busy knowledge seeker role I play in New York. Does that mean that I literally am a different person in these settings? It is clear that my distinctive psychology varies when I'm dealing with my professors, students or family members. Although the same basic distinctive psychology is present, there may be a large enough variation for us to maintain that there are two or more distinct persons. We may consider cases in which there is a drastic change in judgment, for instance Rabi', who is taken out of his family life, in which he is compassionate, caring and a loving father and is forced to join the army and fight overseas. Now he needs to kill people and struggle to stay alive. If there

are different persons in the case of people with multiple personality disorder by virtue of their different distinctive psychologies, then it must be true in Rabi's case. This may be problematic, because many of us do not want to claim that we can be two or more different persons just because of the different social roles we play. Different situations certainly make us act differently, and perhaps even have different kinds of judgments, but to conclude from this that there is more than one person is somewhat counterintuitive, and so if this is true in the case of multiples it may be true for the different social roles that one plays. The best version of such a criticism is given by William James:

'Properly speaking, a man has as many social selves as there are individuals who recognize him and carry an image of him in their mind… But as the individuals who carry the images fall naturally into classes, we may practically say that he has as many different social selves as there are distinct groups of persons about whose opinion he cares. He generally shows a different side of himself to each of these different groups. Many a youth who is demure enough before his parents and teachers, swears and swaggers like a pirate among his 'tough' young friends… Nothing is commoner than to hear people discriminate between their different selves of this sort: "As a man I pity you, but as an official I must show you no mercy; as a politician I regard him as an ally, but as a moralist I loathe him;" etc., etc.'[133]

---

[133] James, William *Principles of Psychology Vol. 1* (New York ,Cosimo Classics, 2007) p.293 -295

This seems to be a legitimate criticism on the part of those who do not believe that there is more than one person in the same body cases like that of the Beauchamp family. But here I would like to go back to our original conditions for remaining the same person over time. Recall that on my account, it is not just any feature of one's distinctive psychology that makes him the same over time, but it is the subset of this psychology that is relevant for morality, which consists in the higher order attitudes. In other words, it is the stand that one has towards his desires that matter for our purposes. This provides us with reasons for thinking that there is a great disanalogy between cases of multiple personality disorder and the different social roles that we play. When I am teaching a class, I do not change the attitude I have towards my desires from the time that I am spending with my friends. If I disapprove of murder in one situation then I will also disapprove of it in the second. And even if there are cases in which it seems like I have different judgments in the different social roles I play, like the examples mentioned by James, in which the person literally has two different judgments of the same action, that person generally endorses only one of those judgments. So when one says, as mentioned above, that he condemns an action as an official, but not as a man, he is really saying that if he were to be doing his job and acting in the position that he fills, these should be the judgments that he gives, but these are not the judgments he does in fact endorse. Since, as a man, this is what he genuinely believes, and this has to do with him, not with doing his job or playing a particular role. But in the case of Christine Beauchamp, each

character actually had different attitudes towards actions and desires. One identified with the desire to smoke while the other found it abhorrent. One enjoyed playing harmful pranks on others while another was appalled by such things. It is clear that in the case of multiple personality patients, it is perfectly plausible to have different persons, in our sense of the term, whereas in the different social roles that one plays it is not the case. This seems to be one of the strongest cases that one can make against synchronic multiple personhood. I have given reason to affirm this possibility. I will deal with another objection to my account in a later section of this chapter.

c. Dennett's Conditions

When discussing the conditions that make one a person at a given time, we talked about Daniel Dennett's six conditions. Wilkes attempts to apply these to patients of multiple personality disorder (like Christine Beauchamp).[134] First, all the different characters seem to be rational, at least in the minimal way we have defined it. Second they must be beings to which intentional predicates are attributable. This condition seems to apply to each of the characters in the Beauchamp family, as is clear from the fact that have had no trouble discussing many of their desires and beliefs and memories. Third they must be beings towards whom you can adopt the intentional stance. That this is possible can be illustrated by Dr. Prince's reactions to the different characters. He treats them differently, talks about them differently, and even feels guilty about neutralizing one of them, since he believes it to be

---

[134] Ibid p. 126-129

“cold-blooded” to end Sally’s life. Fourth, they must be able to reciprocate the treatment mentioned in the third condition. In other words, they must be able to treat other people as persons and recognize intentional states in others, and we know that in the case described, the different characters can attribute intentional states to one another as well as to other people. Sally sometimes recognizes Dr. Prince’s intentions and approves or disapproves of them. They even recognize one another’s intentions, with or without having direct access to them. Fifth, persons must be capable of verbal communication. It is clear from the reports that they give of their actions that each of these different characters is capable of verbal communication. The last of Dennett’s conditions concerns self-consciousness. I have argued that what we need for this self consciousness is being aware of our intentional states and having attitudes towards them. If one is capable of identifying with some of his intentional states and considering others to be alien to him, in other words not mesh at all with the rest of his distinctive psychology, then one is a person. This condition seems to apply to Christine and Sally. Sally recognizes her own intentional states and distinguishes them from those of BI and BIV. In fact, at some point Sally gets access to BIV’s thoughts and is frightened by them and she takes a very strong attitude towards BIV’s desires, distinguishing them from her own, as I mentioned in the previous section. She even claims that her desires are harmless and that she does not have any malicious intentions. This is only one piece of evidence in support of the self-consciousness of each of these characters. In fact, there are many occasions

on which one of the characters explicitly states that she is not to be identified with another. So Sally explicitly denies that she is BI, since they have different attitudes towards each of their desires. This all goes to show that on Dennett's conditions in addition to my condition we have reasons to believe that Sally, BI and BIV are different persons. Wilkes gives five further considerations as to why she believes this to be the case.[135] She does not consider these to be knockdown arguments for multiplicity, but rather believes that they are suggestive, even if inconclusive evidence, in support of it. Let us examine these considerations.

d. Wilkes' five considerations

The first argument that Wilkes discusses concerns the differences in characters and personalities of Sally, BI and BIV. This is very similar to my argument mentioned above. She gives a slightly weak version of the strongest argument in support of multiplicity and mentions the most crucial problem with this argument. This is the problem that I mentioned concerning social roles. She argues that we do not change that much when playing the various social roles. Sally and BI seem to be fundamentally different people, and this difference is too big for us to make it analogous to the different social roles that we play. I have argued in the previous section that it is not only the fact that the different characters are different which proves this multiplicity, but it is rather the types of differences that form the basis of this argument. The different characters have different sets of higher order attitudes, and that is

---

[135] Ibid p. 129-132

what makes them different persons in our sense of the term. So I agree with Wilkes, but believe that her argument is stronger than she recognizes it to be.

A second consideration, and again not a very strong argument in favor of multiplicity, concerns the fact that each of the different characters takes sole charge of the body at a given time. So when Sally is in charge of the body, BI and BIV are not around. There is no struggle among the characters, but one of them is acting while the others are somehow dormant, and it is as though something or someone else has taken over the body. In my account, I do not stress control nor do I give a causal story, but I focus on the attitudes the agents have towards their actions. However, it is important to note that it does matter whose desire led to the action, for if it were BI's desire, then it is BI who had done the action and not Sally. So we can say that there are different agents in the case of the Beauchamp family. They do not approve of one another's desires, and each has actions caused by desires she approves of, and wants to get her own way. That is another reason for believing that there is more than one person in that body.

A third argument that Wilkes mentions concerns the fact that it is not at all obvious which of the different characters is "the real Miss Beauchamp"[136]. Even Dr. Prince himself found it somewhat difficult to determine which character is the original Miss Beauchamp, since at least two of the characters had as legitimate a claim for survival as one another.[137] It seems that this is a

---

[136] Ibid p. 132

[137] Prince p. 185-186

further reason to consider them as distinct persons, since they are independent of one another and neither of them seems to be parasitical off the others, so to speak. So it seems that there are several distinct persons there.

A fourth argument concerns synchronic and not just diachronic multiplicity. When either BI or BIV acts, or is in charge of the body, it seems that Sally is there as well, as though she is observing them. She is present when they act, despite the fact that she is not capable of moving the body at all. Sometimes Sally has more precise memories of BI's actions than BI herself, so it is as though Sally is watching BI and judging her actions. This seems to suggest that there is more than one person present, since there are conflicting attitudes at the same time. While BI is attending Sunday mass and endorsing her desire to pray, Sally is watching her disapprovingly and believes that BI is weak and boring. She would like to be doing other things during the times that BI is in charge. So it seems as thought this is a case in which you have two persons, an agent and an observer. The two not only play different roles, but also take different attitudes towards the actions being done. This, again, seems to suggest that there is more than one person present in these cases.

The fifth and final argument involves an appeal to intuition. Wilkes puts it in terms of Thomas Nagel's argument concerning what it is like to be an X. It must feel different to be Sally than it does to be BI or BIV. Since this is the case, they must be different persons. Since I am not concerned with the

qualitative feel of certain intentional states, but more concerned with the judgments and attitudes towards these states, this argument is not very helpful for my purposes.

As mentioned earlier, none of these arguments are conclusive in proving that there is more than one person instantiated in the same body in the case of patients of multiple personality disorder. However, these arguments seem to suggest that it is more intuitive to consider them that way than to think is only person in every body. I have given further arguments in support of my position. The case of the Beauchamp has helped us shed light on the issue at hand.

This case provides us with reason for believing in synchronic multiple personhood. Each of the persons involved is responsible for her actions. The other person may not be responsible for these actions. Each person is responsible for the actions that the person they are continuous with has done, but is not responsible for the actions the other person had done. So Sally will be responsible for whatever actions the persons she is continuous with had done and were responsible for and Christine will be responsible for whatever actions the persons she is continuous with did and were responsible for. Even if Sally evolved to become a different person, she will remain responsible for whatever the person she is continuous with had done. A more detailed discussion of this account will follow later on, but for now let us turn to another case of multiple personality disorder, although one not as well documented as that of Christine Beachamp.

3. Jonah

The case of Jonah, a 27 year old man, is more recent than that of Christine Beauchamp.[138] Many personalities arose in the same body, and they had very different characters from one another. Jonah is described as *"shy, retiring, sensitive, polite, passive and highly conventional."*[139] One may notice that this is description similar to that of BI discussed above. The researchers decided to designate this character *"the square"*. Another character was Sammy, *"the mediator*" who describes himself as *"purely intellectual, rational and legalistic"* and *"is interested in accumulating knowledge for knowledge's sake and enjoys twisting words and engaging in debates"*[140]*.* A third character, King Young, who was dubbed *"the lover"*, viewed himself as *"pleasure-oriented and as quite a ladies man.. enjoys good times, especially with women, is a glib talker and cannot take "no" for an answer.'*[141] The last character, Usoffa Abdulla was dubbed "*the warrior*", and may be described as *"cold, belligerent and angry person… is generally sullen, silent, occasionally sarcastic and primed to respond aggressively to any threat or challenge"*[142]*.* Although Jonah was not aware of the existence of

---

[138] Ludwig, AM, Brandsma, JM, Wilbur, CB, Bendfeldt, F and DH Jameson. 'The Objective Study of a Multiple Personality' *Archives of General Psychiatry* April 1972, 26 (4) 298-310

[139] Ibid p. 300

[140] Ibid p. 300

[141] Ibid p. 300

[142] Ibid p. 300

these alters, they were all aware of him and what was happening to him, and they were also aware of each other's existence.

It is clear in the case of Jonah and the three other characters that they all have different higher order attitudes, since the descriptions of their personalities gives us good reason for believing that. One of them is angry, and responds aggressively to threats or challenges, whereas another is passive and highly conventional. One of the characters is interested in purely rational activities and only gaining knowledge whereas another is interested in good times and women. Usoffa Abdulla attacked Jonah's wife with a butcher knife, which is something that Jonah would never do, and would never approve of. He also beat up a man and a woman at a bar and was jailed for it, which is also something that is alien to Jonah.

At the end of the study, the writers argue that there are in fact different personalities instantiated in the same body, and not just features of the same person.[143] They deduce this from the variations on a number of exams administered on each of the characters, and deduced that these variations could not have been faked by one person who was engaged in role playing, but were actually given by different people. This case goes further to show that my conclusion discussed above in the case of Christine Beauchamp is intuitive and may be carried through to other similar cases. Now let us turn to how these cases of synchronic multiple personhood are different from trans-temporal cases.

---

[143] Ibid, p. 307-308

4. Difference from Trans-temporal Cases

One can think of the following case: Suppose that instead of developing multiple personality disorder, Christine Beauchamp underwent a lot of changes in character, and her experiences led her to give up many of her old higher order attitudes and adopt new ones. Let us further suppose that she changed so drastically, that she started having a character that closely resembles the original alternate, Sally. Given my account of personhood, it seems clear that in this case we can say that the later person who resembles Sally is different from the earlier one who resembles BI. Since they have different sets of higher attitudes, they must be considered different persons. Is that then the same as the case of MPD discussed above? I believe that the answer to this question is no, because although the later person is different from the earlier one, the later person is responsible for the earlier person's actions. There is a closer connection between the later person and the earlier one in the trans-temporal cases than there is in the case of multiples.

In support of this, I will give an argument that is along the same line as the one mentioned in Chapter 4. In the separate personalities of the Beauchamp 'family', each of these characters is a different agent with different attitudes, concerns and desires. Each of the characters can function independently of the others. So it seems that in the case of BI, BIV and Sally, each of the separate characters is to be held responsible for its own actions and not for the actions of the others. And over time, a later person from the

Beauchamp family will be responsible for actions performed by whichever of the different persons she is continuous with. What is interesting is that in this case there is an actual mergence of two of the characters, BI and BIV, and at the end of the study, Prince tells us that Miss Beauchamp, the result of this mergence, takes responsibility for the actions of both BI and BIV, for "after all, it is always myself"[144]. Since both persons are continuous with that later person, then that later person is responsible for both their actions. Of course I need to clarify that it is her judgment that supports my position, but not her way of putting it. I do not want to say that Miss Beauchamp is the same person as BI and BIV, since she does not share enough of her higher order attitudes with either, but only that she is psychologically continuous with both these persons, and therefore responsible for both their actions. She is a psychological product of both characters, and is therefore responsible for their actions.

However, the case just mentioned is one in which one of the persons, undergoes radical changes and so comes to have a character that closely resembles a different member of the Beauchamp family, and so the later person is still responsible for the actions of the earlier person. The later person's psychology and higher order attitudes are caused by the earlier person's psychology. So in some sense the earlier person brought life to this later person, and so the later person is responsible for the earlier one's actions. The later person came about from the earlier person, since the

---

[144] Prince, p. 525

distinctive psychology of this later person is the causal product of the earlier person's psychology. What makes something a particular person is the fact that he has a particular distinctive psychology, and if that distinctive psychology was caused by an earlier person's distinctive psychology, then this later person owes his particularity to the earlier person. It is the earlier person and his actions that caused the later persons to exist. This is the same argument as the one I mentioned in Chapter 4, and should apply to this case as well. As long as that person was continuous with the one who had originally performed the action, they are responsible for it. It also follows that even if the later person endorses the desires that led to actions that Sally had done, and not those of Christine, she would still be responsible for Christine's actions, because she is continuous with her. What counts for responsibility is continuing to be a person, and not being the same person.

One objection that might arise to this argument concerns the causal story. Surely it must be BI's psychology and her past experiences that caused the creation of both Sally and BIV. After all, these characters emerged because of the original person's inability to deal with some traumatic experiences. The DSM IV discusses the reasons for the emergence of multiple personalities in the same body, and among them is the original person's inability to deal with some traumatic experiences.[145] So, on my argument these persons, i.e. Sally and BIV. should be held responsible for BI's actions. On my account they should be responsible for the past person's

---

[145] Diagnostic and Statistical Manual for Mental Disorders IV-TR Fourth Edition, p. 482-487

action. Each of these two persons is responsible for what the person they are continuous with had done, and in this case, they are both continuous with the original person, and hence responsible for her actions. However, although they both remain responsible for the original person's actions, in other words the actions of the person before the split whom they are both continuous with, they are not responsible for one another's actions after the split. So both Sally and BIV are responsible for the original Christine Beauchamp's actions, but Sally is not responsible for BIV's actions, nor is the latter responsible for the former's actions

One Tricky Case

One new and interesting case is a slight variation of one that was mentioned in the previous chapter. Suppose that the BI changes radically over time, and her character gradually becomes very similar to that of Sally while Sally continues to exist normally, without too many changes in her character. So at this later time, we have the two persons BI and Sally, but now they both share a large part of their distinctive psychologies and higher level attitudes. Can we now say that Sally and BI merge to become the same person? If so, does that mean that Sally is now responsible for the earlier actions of persons continuous with BI? And if so, then it seems to contradict with what I had mentioned earlier, namely that each of the characters is responsible for actions of persons that they are continuous with.

There are two possible ways in which this may occur. The first way is one in which the characters BI and Sally really do come to have the same

distinctive psychology and memories, in which case I may safely say that this new character, call her Chally, would be responsible for the actions of all the persons she is continuous with. This is essentially the account I gave concerning the mergence of BI and BIV to form what Dr. Prince calls “the original Miss Beauchamp”. This case is not problematic for my account.

Another perhaps more problematic way is one in which both the later BI and Sally share most of their higher order attitudes, but do not share a distinctive psychology. So although Sally remembers all of BI’s previous actions, BI has no recollection of Sally’s actions. Here there may be an asymmetry between the two characters, which, in terms of higher order attitudes are very similar, but are not the same in the rest of their distinctive psychology. They come to have or share similar higher order attitudes, but they do not have the same distinctive psychology. So there is no Chally in this case, since Chally is one person who happens to be the mergence of these two characters. In this case, Sally and Christine did not merge to become one person, because some part of their distinctive psychology is not shared. It is higher order attitudes that makes one a person, and the same particular set of higher order attitudes that makes one the same person, but if there is a large enough set of lower level states that the two do not share, then they are not the same person. Since they come to share some of their of their psychology, they still do not have the same distinctive psychology like in the case mentioned above. One reason for this can be stated in terms of higher order attitudes and another cannot. Since many of their lower level states are

different, this may be reflected in their higher order attitudes, for Sally recognizes some of her lower order states and takes attitudes towards them whereas BI has no such access to these states.

These two cases mentioned above are not the same, because in the former case there is a fusion of Sally and Christine, and they formed this new person Chally, which is the same situation as the mergence of BI and BIV to make the original Miss Beauchamp. In the latter case, however, Christine only became someone who has a character that is similar to Sally's, and there is, in some sense, simultaneous endurance of twins, where they share some of their psychology, but do not have the same psychology. In the first case they merge and are both continuous with Chally, whereas in the second case they remain distinct. In the first case, Chally is continuous with both BI and Sally and in the second case, Sally, is continuous only with herself, as is BI.

The later BI cannot share too many higher order attitudes with Sally if they do not share most of their lower order states. Another way of looking at it suggests that these lower states interact with higher order attitudes in a causal way that is distinct in Sally from the way it is in BI. These two therefore have different distinctive psychologies, although they do share some of the higher order attitudes. This case is not as difficult as it seemed at first. Memory, as I mentioned in Chapter 4, is only important insofar as it plays a causal role in shaping and changing higher order attitudes. Differences in memories may therefore create differences in what matters to us, namely the higher order attitudes. So Sally and the later BI cannot be the same person,

and each of them is responsible only for the actions of persons that they are continuous with.

5. Two Majors Objection

a. Cases of MPD Same as Normal People

One objection concerns the fact that you should say the same thing about cases of multiple personality disorder as you should about normal people. In other words, the fact that a human being has multiple personality disorder does not generate any new or interesting metaphysical truths. Rather, we should give a description of this case that is similar to the description we give of other normal humans. In other words, there is no distinction in terms of personhood between cases of multiple personality disorder and other humans, and whatever you say of one you ought to say to the other. This objection can be stated in two ways. First, we can say that there is synchronic multiple personhood in the case of multiples, but the same goes for all humans. So since we believe that there are two or more distinct persons in the case of the Beauchamp family, so too we should say of ourselves, that there are two or more persons incarnated in us. So Dr. Prince is also two or more distinct persons the same way his patient is. Second, we can say that there is only one person in both cases. So just as Dr. Prince is to be counted as one person, so too should Christine Beauchamp. There are no cases in which you have more than one person incarnated in the same body. I will deal with each of these versions of the objection next.

Many philosophers and psychologists have argued to the effect that each human being is not only one person but many different ones. This, in other words, is not something particular to patients of multiple personality disorder, but is true of all of us. The major difference between the multiples and other people is that the different characters are more extreme in the cases of multiples than in the rest of us. Adam Crabtree argues that it is naïve to think only one agent per body, not just in the case of multiples but in all humans. In every human being there is more than one self and it is impossible for us to just find out what the real self is.[146] Puccetti discusses many experiments involving severing the corpus callosum, which reveal that the two hemispheres of the brain function in different ways. He uses this to argue that there isn't one unified consciousness but many consciousnesses, which means there are many selves.[147]

One can see how this objection is a more radical version of the one we discussed above concerning the different social roles that one plays. So it seems natural that I adopt a similar response to the one I gave to the other objection. Even if Crabtree and Puccetti are right, the most they show is that some mental phenomena operate independently of others. Recall that our concern is with the kinds of beings that have higher order attitudes, which means they are capable of recognizing and evaluating different intentional

---

[146] Crabtree, A. *Multiple Man: Explorations in Possession and Multiple Personality* (Toronto, Praeger Publication, 1985)

[147] Puccetti, R. 'Brain Bisection and Personal Identity' *British Journal for the Philosophy of Science*, vol. 2, pt. 4, pp. 339-355, December 1973

states. It is doubtful that most human beings have more than one set of higher order attitudes, and that they have one part approve and another part disapprove of the same intentional state. So even if proponents of the multiple man view are right, this has no bearing on personhood in the sense of the term which is relevant for moral responsibility. Moral characters are made up of a collection of interacting higher order attitudes, and people do not normally have different sets of higher order attitudes. Puccetti does not give such cases. That is why cases of multiple personality disorder and interesting and crucial for our purposes. The different characters actually identify with different desires, and they usually have different desires as well. So this version of the response does not work against my account.

The second version of this objection involves claiming that in both normal cases and in cases of multiple personality disorder there is only one person. Most of us seem to think that people with multiple personality disorder are disturbed individuals with a strange psychological condition. Generally the genesis of these different personalities is through some traumatic experience or experiences, and it may seem extreme to consider these characters to be distinct persons. It seems counterintuitive to hold a personality responsible for its actions, if it is not in fact a distinct personality, but the original person's way of coping with troubles. Psychologists have come up with explanations of why these characters come into existence. Even Prince himself discusses how all these dissociations turn into a distinct personality.[148] When a person does not

---

[148] Prince p. 20-52

associate with some psychological state that is in fact his, psychologists talk about dissociation. In these strange cases of multiple personality disorder what happens is that there are so many of these dissociated psychological states that the person ends up dissociating a whole personality. So in the case of Christine Beauchamp, all the different subconscious desires that she had but did not associate with came together to form an apparently totally distinct personality, that of Sally. Because of this, many people might argue that it is unfair to characterize these cases as literally involving “totally distinct personalities”, and instead of saying that there are two distinct persons in one human being, we should say that there is just one person with a psychological disorder. Some may even argue that there is no person at all in our sense of the term, because the disordered personality does not qualify as the sort of being that can be held responsible for its actions. This objection I will deal with in the next section, but now I will tackle the one concerning the fact that there is only one person in our sense of the term. I consider this to be a major objection to my account.

Consider the case that we discussed in the previous section, namely the one where Christine gradually turns into Sally. It seems perfectly plausible to say that these two characters are separate persons on my account. Consider even further that Christine and Sally were instantiated in two different bodies and they that had these sets of higher order attitudes. In this case there seems to every reason to consider them to be two separate persons who can be held responsible for their own actions. One may object

here by saying that this case is completely different from the case of the multiples, because in the case of the multiples the different persons are incarnated in the same body, and the fact that they arose is altogether the result of a psychological condition. This is a clear instance of the genetic fallacy, because the only argument that can be appealed to concerns how these persons came about, and not the distinctive features of these persons. That may be true, but if someone were to object that way, they would have to provide us with a reason why the same human being cannot be two different persons, and why this unnatural development somehow exempts the different characters, who otherwise would have been persons, from being persons in this case. I have argued in Chapter 4 that what makes two persons different from one another is their distinctive psychology, and more precisely their distinct sets of higher order attitudes. Since in the cases of multiples they do have different higher order attitudes, then there is good reason to say that the distinct characters are different persons. The onus here is on the objector to provide us with good reasons for dismissing this hypothesis given that these cases satisfy all our conditions of personhood. Relying merely on the fact that they are in the same body obviously begs the question against this account. There is nothing about being in the same body that automatically implies that we only have one person. When examined closely, this objection seems to be based on a question-begging assumption. But someone might take it a step further and argue for the fact that in these cases there is no person at all in our sense of the term.

b. Not Persons

Another objection to my defense of synchronic multiple personhood goes as follows[149]: People who believe that in cases of multiple personality disorder there are different persons incarnated in the same body would suggest that each person is not responsible for the done by other persons, but are nevertheless responsible for their own actions (and those of the persons they are continuous with). This does not sound like the way we want to deal with such cases, for it is not the fact that it is impractical to punish one personality for the action of another that stops us from doing so, but the fact that we seem to think the person had not been in control of his actions, even the personality that was in charge. It is probably beneficial here to draw an analogy to drunkenness. When someone does a certain action when they are drunk, we don't hold him responsible for the action he did but we would blame them for getting drunk knowing that such a thing might happen. So if someone were driving while intoxicated, we tend to think that she should not have drank so much knowing that she was going to drive eventually, but we do not blame her for getting behind the wheel and driving when she was already intoxicated. Similarly, in cases of multiple personality disorder, we may not want to blame the personality for doing something that it was not really in control of, or an uncontrollable impulse. Perhaps we would want to blame the original person for dissociating and in effect creating the alter, but

---

[149] One such argument can be found in 'Identity, Control and Responsibility: the case of Dissociative Identity Disorder' by Jeanette Kennett and Steve Matthews, *Philosophical Psychology,* vol. 15, no. 4 2002

that also seems to be counterintuitive, for we would be condemning someone for something they had no control over. It is unjust to blame one for not being able to cope with traumatic events except by dissociating. This is why none of the alters should be considered responsible, for they are all interconnected. There is no person in our sense of the term, just a human being with a psychological disorder that needs to be cured.

I think that this is a fair objection, much stronger than the ones mentioned above (that there is either one person in cases of multiple personality disorder or that there is synchronic multiple personhood in normal cases just like in the multiple cases). We are tempted to say that subjects of multiple personality disorder are not sane, and therefore are excused from their behavior, since they have very unorthodox ways of dealing with their troubles. But the fact that this way of coping is unorthodox is what inclines us to lump it under the category of insanity, which is the default classification for things that we find difficult to comprehend. The truth of the matter is that these different characters satisfy our conditions of personhood. If one were to dismiss them right off the bat as being insane, he would be overlooking the moral dimension of these different characters, who have commitments and beliefs just like any of us. They are capable of holding themselves responsible for actions, they boast about achievements, they hold one another responsible for actions, and they even have different reactions to the same situations. As I mentioned in the earlier chapter, had these characters been instantiated in different bodies rather than the same one, we would have

easily considered them different persons. This is why it seems strange to rule out synchronic multiple personhood based on insanity, since the case of the "Christine Beauchamp family" gave us many reasons to differentiate the component characters from one another and consider all of them sane. These characters satisfied all the conditions for being a person, and that is why we should consider them to be such. None of the members of the Beauchamp family seem to be individually insane. Their behavior is generally rational and consistent with their desires and attitudes, and that is why I see no reason to exclude them from personhood and responsibility.

Conclusion

Cases of multiple personality disorder have puzzled many philosophers who have discussed personal identity. It is difficult to see what one ought to say about these cases. In this chapter, I have applied my conditions to these cases, and I believe I have given answers to questions concerning synchronic multiple personhood. This application also happens to fit with our intuitions about these matters. Since what makes someone a distinct person is the unique set of higher order attitudes that they have, then sufferers of multiple personality disorder are distinct persons in our sense of the term. One character cannot be responsible for another's actions, unless this character is continuous with the earlier one. In the case of multiples we do not have the overlap of psychologies that justifies one's being responsible for the other's actions. BI was the shy, passive, reserved but mature character, while Sally was the playful immature child. BIV was the mature

character who was more active and less patient than BI and did not tolerate things that BI was able to tolerate. Prince thought the fusion of BI and BIV would create the Real Christine Beauchamp**.** Even after fusion, as in the case of Christine Beauchamp, the later fused person would be responsible for the actions of the persons she is continuous with. So since Christine Beauchamp is the fusion of BI and BIV, she is responsible for their actions, but not of those of Sally. It is interesting to see cases of multiple personality disorder and assess whether they would allow the possibility of synchronic multiple personhood. Cases usually discussed in the literature are often science fictional, and no more than ideas to play around with. The case of multiple personality disorder is different. It is a discussion of phenomena that occur in real life and that shed light on the issue at hand.

Chapter 6 Tricky Cases and Possible Problems

In this last section, I will try to tackle possible problems with my account. I will also deal with tricky cases and see how the application of my theory would give us answers to these questions. Also, I will give some more examples and thought experiments that support my account. Some of these cases have been discussed in the literature, while others have been overlooked. I will argue that other theories fail to give satisfactory answers to some major objections while my account succeeds.

1. Higher level endorsement

Although Harry Frankfurt presents a view of personhood that focuses on second order volitions, he never actually gives a theory of personal identity over time. I do give such an account. Since higher order attitudes play an essential part in my accounts of both personhood and responsibility, it is natural to examine cases in which some higher level attitudes are variable. In Chapter 4 I discussed cases in which the person who had originally done a certain action then changed over time to become a different person with a different set of higher level attitudes. As I argued earlier, one may easily imagine a case in which a person performs some action which he higher-order endorsed at the time he performed it, but changes later on and stops endorsing it. We should say that the later person is still responsible for that action if he is psychologically continuous with the person who did it. Also, even if he changed radically, and no longer had many of the higher order attitudes that the person he is continuous with had, he remains responsible

for the actions that this earlier person had done. Continuity of personhood is what matters for responsibility, and not personal identity.

a. Appearing Higher Order Endorsement

I would now like to turn our attention to a new case in which the person who originally acted did not have a higher order endorsement of the desire that led to his action. In fact, this person disapproved of the desire that led to the action, but the action was caused by a lower-order desire that he did not identify with and that he just could not resist. So the attitude that is necessary for holding him responsible for that action was absent. Later on, the person changed and developed a higher order endorsement of the desire that led to the action and identified with that desire. The question arises whether in these circumstances the later person is responsible for the action, whereas the earlier person who did it was not. According to my account, the person who acted was never responsible for his action, and even though the later person endorses it, he is not responsible for it either. We might judge his character according to this endorsement but we would not hold him responsible for doing it. We may perhaps be appalled by such an endorsement if we considered the original action to be immoral, for example rape, but we cannot hold that person responsible for this particular action. It would be like being appalled by someone who endorses rape, although he has never raped anyone.

One can object to this by saying the following: Since it is the existence of this higher order endorsement that makes a person responsible for an

action at the time he did it, then it is reasonable to argue that developing this higher order endorsement later on makes him responsible for that action. In other words, if the very element that makes one responsible is present, then his excusing condition disappears, and he becomes responsible for that action. This is a serious objection that I will deal with next.

On my account, being responsible for an action at a given time is dependant on the presence of a higher-order endorsement of that action. However, being responsible for that action at a later time does not require the maintenance of this endorsement, but is determined by psychological continuity. As long as the distinctive psychology of the later person is causally continuous with the distinctive psychology of the earlier person, the later person remains responsible for the earlier person's actions. Responsibility is not tied to the appearance and disappearance of particular higher order endorsements which are the bases of personal identity. These endorsements are important at the time the action is done because they are part of the person's moral character, making any action in line with them his own. Since the feelings towards the action are positive, then the person is acknowledging the fact that they had done it. He claims responsibility for the action while he is doing it, which demonstrates that it is really natural for him, and fits well with his character. However, he cannot make it his own at a later time. Although the connection between any person and other persons that he is continuous with is a strong one, he cannot retroactively make an action his own, or a result of his character. That is, of course, different from one

realizing at a later time that he had a higher order endorsement of the action at the time he committed it. In this case he would be responsible, although he did not realize it until a later time. So if someone really did not have a higher order endorsement in line with his action, then neither he, nor anyone he is continuous with, can be held responsible for that action. The later appearance of a higher order endorsement does not change that fact.

b. Re-appearing Higher Order Endorsement

One objection against any account of personhood that relies on character is based on the idea that character traits, which include higher order endorsements on my account, can change and therefore cannot be the bases of continued responsibility.[150] The following example illustrates precisely this point.

Let us consider a case in which the agent had a higher order endorsement of the desire that led to his action at the time it was done. Later on, he ceased to endorse this desire. For instance, consider Michael, a willing drug addict who stole in order to buy his drugs. Michael had a higher order endorsement of the desire that led to his theft. Later on, Michael underwent a religious conversion. He stopped endorsing many of the desires that he used to have (or maybe still has, given the fact that it is difficult to stop being a drug addict suddenly). We may say that the religious man is a different person from the drug addict, since many of his higher order attitudes have changed. Yet after a while, he went back to identifying with this desire.

---

[150] Perry, J. 'The Importance of Being Identical' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 67-90

So after being extremely religious for a while, something terrible happened to him and he lost faith, sliding back to his old habits of drug addiction. It may be argued that he went back to being the same person as the drug addict. So Michael started off identifying with the desire that led to the action; later on rejected it as one that he no longer endorses; but at an even later time came to endorse it again. We might have thought that in the middle time interval when he stopped endorsing the desire, he changed and became a different person from the one who had committed the action. Yet if we were to say that he became a different person and therefore was not responsible anymore, it seems counterintuitive to say that he somehow became responsible again when he went back to being the same person.

However, my account deals with this problem in the following way. All three persons at the three different times are responsible for the action. All that is needed for responsibility is psychological continuity, and at the three different times the persons are psychologically continuous with one another. This case illustrates the claim that having the same character (that is being the same person) is not the basis of remaining responsible for an action. On my account, the original person was replaced by a different person, but later on that new person was replaced by the original person (because he is both psychologically continuous with and has a similar set of higher order attitudes to the earlier person). None of these persons was ever exempted from responsibility for the action. The changes in higher order attitudes resulted in changes in characters, and therefore in personhood, and they allowed new

persons to replace older ones, but they did not affect responsibility. On my account, the conditions of personal identity over time are different from the conditions of responsibility over time, and thus varying higher order attitudes do not affect responsibility. It seems intuitive to say that the earlier Michael and the later Michael are different persons. The only counterintuitive step is exempting the religious man from responsibility because of the change in his attitudes, but on my account the religious man is not exempted. So although the person's attitudes towards an action at the time he performs it determine whether he is responsible for it or not, keeping that attitude or changing it has nothing to do with whether he will remain responsible for it at a later time. A person may be responsible for another person's actions even if he has no favorable attitude towards these actions. Psychological continuity is all we need for responsibility.

2. Memory

John Locke (and Leibniz[151] before him) focused on memory as the essential condition for both remaining the same person over time and remaining responsible for an action over time. On my account, memory is essential only insofar as it plays some role in the development of the higher order attitudes that a person has. Now let us turn to some difficult cases involving memory.

a. Responsibility and Amnesia

---

[151] Leibniz, Gottfried "Discourse on Metaphysics" Tr, R.S. Woolhouse and Richard Francks, *G.W. Leibniz Philosophical Texts* p. 86 (New York, Oxford University Press, 1998)

Let us examine a case in which someone experiences complete amnesia. This is a case in which the slate has been wiped clean, and the new person is starting fresh, like a newborn baby. In this case, we may say that there was discontinuity in the development of these later higher order attitudes, and therefore the later person is no longer responsible for the actions that the earlier person had done. This is one of the rare cases in which a person may actually be exempted from an action performed by an earlier person instantiated in the same human body. It is only in such cases of psychological discontinuity that exemptions are possible. This is an example that defenders of the memory account rely on since it seems intuitive to exempt a person from responsibility in this case. However, it is not the fact that the later person does not remember that exempts him from responsibility; it is the psychological discontinuity that exempts him from it. I hope the next few cases illustrate this point. In other words, to draw Lockean conclusions from such cases begs the question against my account. Lockeans say that memory is necessary for continued responsibility because, absent memory, there is no personal identity. They take for granted that continuing identity is necessary for responsibility, and explain the failure of responsibility via failure of identity. My view is that there is no responsibility because there is no psychological continuity between the earlier and later persons. I am therefore not obliged to explain why personal identity is destroyed in cases of complete amnesia but not in cases of gradual change of character, since I do not think

identity is preserved in either case. So these cases are not a problem to either my account of personal identity or responsibility.

On the other hand in a case in which a person has partial amnesia, and this person's current higher order attitudes are caused by the earlier person's psychology, we can say that the later person is responsible for the earlier person's action, although he does not recall anything about that action. As long as the person was responsible for the action at the time he did it, and the later person is psychologically continuous with that earlier one, the later person remains responsible for the action of the earlier person. This is obviously a controversial claim, and many people might consider it counterintuitive to hold someone responsible for an action they don't remember doing, but I maintain that we should (and do) continue to hold them responsible. Many psychologists discuss how traumatic experiences are often repressed by the people who had them.[152] As a psychological reaction to such traumas, people use repression as a defense mechanism so they do not have to remember actions that they are not proud of, or that they do not want to remember. But this does not seem like a good reason to exempt these people from responsibility.[153] The mere fact that one does not remember that action does not exempt him from being responsible for that action. This is where the memory theorists go wrong. They are committed to the claim that

---

[152] For a good discussion of this, see Crabtree, A. *Multiple Man: Explorations in Possession and Multiple Personality* (Toronto, Praeger Publication, 1985)

[153] Analogously, we wouldn't exempt someone from responsibility if they hired a therapist to remove their memories of their bad actions.

the later person cannot be held responsible for an action done by a different person, and since he does not remember having done it then he is a different person from the one who did it, and therefore is not responsible for it.

We may even consider a more ordinary case. Suppose that, when he is 12 years old, Hazem plays a cruel trick on Tarek that publicly humiliates him. At the time, Hazem endorses this impulse to play this trick on Tarek. Tarek ends up being picked on throughout his school years, and this results in his loss of self-esteem which makes him depressed. Many years later, say when Hazem is thirty, he has completely forgotten the incident. But Tarek was severely affected by this and his life changes. Hazem then finds out that Tarek is now a bum on skid row, mostly because of the trick that Hazem does not remember playing. Still, Hazem will be responsible for the trick, and may feel to some extent responsible for Tarek's subsequent fate. So memory is not necessary for responsibility. The reason Hazem is responsible is not the tacit premise that he is the same person, but the premise that the person he is now (same or different) is a natural, continuous extension of the cruel child, whether he remembers the trick or not.

Memory theorists do not rely on memory insofar as it plays a role in shaping the later person's higher order attitudes, but focus on it for its supposed role in personal identity and responsibility. There is no role that memory plays in our normative judgments about responsibility, and that makes it largely irrelevant to rely solely on this feature in one's account of

continued responsibility over time. The next example, I hope, will illustrate this point even further by bringing up odd cases of people's ability to remember.

b. Recovered Memory

One interesting case that supports my view is the following. Consider a case in which a partial amnesiac forgets that he had committed a certain crime. However, after undergoing therapy, he regains his memory. It seems strange for someone to say that at the point at which he had forgotten what he had done he became exempt from responsibility, but that he starts being responsible again after he remembers. This, of course, is what the memory theorist is committed to, unless he argues that in the intermediate period he does really remember but fails to acknowledge that he remembers, or else comes up with a memory-based criterion for personal identity that somehow allows for persistence of identity over memory loss and amnesia. That, however, is a strange way of casting the story, because memory implies the ability to access the information from the past, so if he is unable to access this information it seems strange to say that he remembers in any sense. If, on the other hand, memory theorists were to claim that memory is crucial only insofar as it affects the rest of the person's psychology, as I do, then they can claim that repressed memories still play a role and therefore can be considered a part of the distinctive psychology of the person. My account deals with this case more easily, because it would not matter whether he remembers what he had done or not, but whether he continued to be a person throughout. In other words, whether it is a robust memory or a

repressed one that he is able to access at a later date, this memory still plays the same role, which has to do with how it affects that person's distinctive psychology. This is perhaps the most difficult case dealing with memory, and I believe that it is very problematic for an account of responsibility that relies solely on memory and no other psychological states.

3. Physical Continuity Views: Why They Miss the Question

In previous sections of this work I mentioned Peter Unger's physical continuity view, in which he argues that physical continuity is crucial for personal identity only insofar as it supports core psychology. I believe that the spirit of his claim is true. Although I believe that the continuity of core psychology is essential for remaining responsible for an act over time, I believe that it is not sufficient for being the same person over time. I agree that physical continuity is important only insofar as it supports the cognitive capacities and the distinctive psychology of the persons in question. It is important to note that Unger believes that personal identity is what matters for survival, which involves concern for the future, whereas I am interested in being responsible for some actions, which involves the past. Both of us, however, consider continuity of core psychology to be essential to our concern. Although I do not believe that we need personal identity to have continuing responsibility, nor do we need continuing responsibility to have personal identity, he believes that we need concern for the future self to have personal identity. He defines the future self in terms of what we have concern

for, and so defines personal identity in terms of these as well. This is the crucial disagreement between us.

Unger believes that continuity of core psychology is what we have concern for in the future despite changes in distinctive psychology. This continuity is also what determines personal identity. So according to Unger, continuity of core psychology, what we have concern for in the future and personal identity are all the same. On my view however, continuity of core psychology is what we need for responsibility (again despite changes in distinctive psychology). So that is parallel to Unger's view. However, what differs is our view of personal identity, which I believe is not the same as these two. Continuity of core psychology makes one responsible for a past action, but it does not make him the same person as the one who performed that past action. This is the crucial disagreement between Unger and me. I believe that the problem with Unger's view is that he makes the same assumptions as everyone else about personal identity and what we have concern for. Granted that my concern is different from his, namely mine has to do with past actions and responsibility for these action and his has to do with the future and whether we have concern for a future person, I believe that personal identity does not determine continuing responsibility. In other words, one may be responsible for an action that a different earlier person had committed, as long as these two persons have a continuous core psychology.

However, other proponents of physical continuity theories seem to be very misguided concerning what the question is. One theory based on physical continuity is presented by Eric Olson.[154] Olson believes that moral responsibility is not the main concern when we talk about personal identity. He argues for the view that psychology has nothing to do with personal identity, and that since we are human beings, then it must be that biological continuity is what is needed for personal identity, and not psychological continuity. What is interesting in Olson's view, however, is that he discusses the issue of moral responsibility and its relevance to personal identity, and his view of moral responsibility is quite similar to mine. He writes:

"Someone is now responsible for something he did earlier if and only if he is morally competent now and he was morally competent at that earlier time, and he is now psychologically continuous with himself as he was then."[155]

It seems clear here how his view is very similar to mine, except for the fact that he takes personal identity to be necessary (although not sufficient) for remaining responsible for an act over time. This goes further to show the intuitive appeal of my account of moral responsibility over time, although he falls into the same trap as others in considering personal identity to be at least necessary for moral responsibility over time. This is the claim I have denied in

---

[154] Olson, Eric *The Human Animal* (New York , Oxford University Press, 1997)

[155] Olson, Eric *The Human Animal* p. 62

this dissertation. His view of personal identity is therefore divorced from psychological concerns, and is based purely on biological concerns.

However, it seems here that instead of answering our original question about personal identity over time, Olson is answering a different question, that of human identity. In the first chapter of this dissertation, I made the same distinction that Locke had made between personal identity and human identity, and said that the main concern of personal identity is different from that of human identity. Human identity over time may be of some interest to biologists and physicians, and there may be some questions concerning it that are interesting in their own right, it has nothing to do with our questions concerning the core issue of personal identity.

The concern behind the issue of personal identity is a normative one, so understanding what it means to talk about persons in the forensic sense of the term has to involve something normative. I'm not sure why proponents of this view base so much on physical continuity unless they believe that it somehow tells us something about the psychological states of the agents in question, unless, like Olson, they are concerned with human identity. If the motivation behind their account is that you cannot have the same intentional states unless you have the same physical being, then physical continuity becomes, as Unger argues, only derivatively important. But if they, like Olson, believe that it is only biological continuity, and not psychological continuity, that gives us personal identity only because we are human animals, then they are dealing with a different issue than the one we are concerned with. The

crucial elements for our concerns are the psychological states, and particularly the higher order psychological states, and it is these alone that constitute the types of beings that can be held responsible for their actions. Whatever needs to be present physically for our psychology and therefore for responsibility to kick in is of course necessary for responsibility by definition, but only because it is necessary for our psychology. It wouldn't be a condition of responsibility for beings that are psychologically constituted in a different way. We are taking the intentional stance, and talking about physical continuity seems to involve taking a different stance. It becomes difficult to talk about normative worries when we are discussing the firing of neurons. This is precisely what I agree with Unger on. Although that is what is going on biologically when someone is entertaining a higher order attitude, it is completely irrelevant for our normative discussion. Making a relevant connection between physical continuity and continued personhood is a very difficult task once we have underlined responsibility as what is crucial to us. What we are concerned with are attitudes that people have towards desires and actions of themselves and of others, and unless one is a scientist knows how that translates into physical activity, there is no point discussing fine grained details of the physical aspects of it. The level of discourse is just a different one and the physical instantiation of these higher order attitudes is only important because of these attitudes themselves. Here is another thing I agree with Unger on.

One way of cashing out this criticism is the following. Physical continuity theories seem to be conceptually misguided. Since physical continuity plays no conceptual part in our story of personal identity or responsibility over time, then it is not conceptually necessary for the persistence of either persons or responsibility over time. It is only a contingent condition. What physical continuity theories are concentrating on is a factor that may be de facto necessary for the sort of continuity required for responsibility but it is not conceptually necessary for personal identity or for responsibility.

This is why I believe that physical continuity theories are misguided. That is not to say that they are wrong on their own grounds, for it may be that physical continuity is crucial for remaining a person over time, but they are only crucial because of their relationship to the higher order attitudes. Recall that these theories are different from Unger's view, and some of these theorists might even consider Unger to be a proponent of psychological continuity.[156] These theorists emphasize biological or physical continuity with no concern for psychology at all, and that distinguishes them from Unger. Although misguided, these theories are not necessarily wrong, nor do they conflict with my account.

Concluding Remarks

I have argued in this work that being the same person as one who performed an action is a sufficient but not a necessary condition for being

---

[156] Olson, Eric, http://plato.stanford.edu/entries/identity-personal/

responsible for the action. I have given an account of personhood, and how it changes over time. I have also given an account of responsibility, and how one remains responsible for an action over time. I have shown that these two do not coincide, and I have given some examples and thought experiments in the last two chapters that make my account more appealing.

I would like to end this by talking about the motivation behind my accounts of personal identity and responsibility. I believe that in philosophy, it is always crucial to answer questions concerning why an issue is of any importance and why one's account is anything more than just a technical improvement over earlier accounts. When John Locke started the debate on personal identity in the "forensic" sense of the term, he was interested in the issue of moral responsibility, and believed that giving an account of personhood and personal identity over time would shed light on that interesting core issue. Many philosophers have since picked up the debate, but have gotten lost in the many difficult thought experiments that have come up. They have lost sight of what I believe was the primary issue at hand.

I believe that the issue of personal identity is important largely because it sheds light on the issue of continued responsibility. However, it seems as though the concept of a person is not the concept to lean on when determining whether one is responsible for an action. We are more readily willing to accept that one has changed to become a different person than we are to exempt that person from responsibility for past actions. Contrary to the implicit assumption of most philosophers, I believe that this readiness is

justified, because if we suspend the assumption that philosophers have made that being the same person over time is both necessary and sufficient for remaining responsible for an action, we will avoid many of the problems that they face. It is true that having an account of personhood is crucial for giving an account of continued responsibility because after all, defining what types of beings can be held responsible for their actions must tell us something about the nature of responsibility. But continuing to be the same person and remaining responsible for an action are not the same. We want conditions of remaining responsible over time that are much more strict and stringent than anything we are capable of saying about personal identity over time for the forensic notion of a person.

One may ask the following question: Why is it that most philosophers fell into this fault and failed to recognize this distinction? Assuming they were interested in the question of moral responsibility, I believe that they made the same mistake that Locke originally made. After defining the term “person” as the kind of being that can be held responsible for its actions, Locke assumed that only beings that are personally identical will continue to be responsible for that action. Since personhood is defined in terms of responsibility (because a person is the kind of being that can be held responsible for its actions) then it must be that responsibility should be determined in terms of personal identity. Since they defined the term solely for the sake of determining responsibility, then it seems natural for them to think the two go hand in hand. Locke introduced the term in order to help us with

responsibility, but failed to recognize that this term does not do all the work. You do need the term, and you do need a concept of personhood and personal identity to be able to shed light on the issue of continued responsibility, but it is not the only thing you need. This concept is helpful in determining something of concern to us, namely responsibility, but that is not the same as saying that this concept gives us a necessary and sufficient condition for determining responsibility. This is where I believe philosophers have gone wrong in the issue of personal identity and continued responsibility.

The issue of personal identity is still one that is interesting for us. We are concerned with whether Zein the drug addict is the same person as Zein the charitable soccer mom. This is because we are interested in her character, interested in describing her, and interested in understanding her past behaviors and predicting future ones. Personal identity is a matter that is of interest for us, but it is not for the reasons philosophers have thought. Persons are determined by their distinctive psychology, and if there is a large enough change in a human being's distinctive psychology, then that human being becomes a different person from the one he was before that change. This is an interesting claim in its own right. But as interesting as this claim may be, it is not what gives us continued responsibility. The fact that Zein has become a loving mother and a helpful member of the community tells me something about that person and that person's character, but this does not

exempt her from responsibility for the past actions of Zein the drug addict. These two issues are separate and ought to be treated as such.

I have shown in this work that if we separate personal identity over time from continued responsibility, then we will get a much clearer picture of what is going on. The reason many people get into confusions when dealing with this problem is that they fail to make this separation. The bottom line is this: you can change and become a different person by changing your character and higher order attitudes, but that does not exempt you from moral responsibility for actions that the persons you are continuous with had done. If this is the picture we have, then the issue of personal identity over time will not give us the headaches that philosophers always complain about.

Bibliography:

Adler, J. 'Constrained Belief and Reactive Attitudes' *Philosophy and Phenomenological Research*, Vol. 57, No. 4 (Dec., 1997), pp. 891-905

Allen, N.J. 'The Category of the Person: a Reading of Mauss's Last Essay' from Carithers, M. and S. Collins and S. Lukes (ed.) *The Category of the person* (New York, Cambridge University Press, 1985) p. 26-45

Allison, H. 'Locke's Theory of Personal Identity: A Re-examination' *Journal of the History of Ideas*, vol. 2, pt. 7, pp. 41-58, January-March 1966

Anscombe, E. 'The First Person' from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 71-82

Aristotle, *Nicomachean Ethics* Tr. T. Irwin (Indianapolis, Hackett Publishing Company, 1999)

Aristotle, *De Anima* Tr. H. Lawson-Tangred (London, Penguin Books, 1986)

Arpaly, N. *Unprincipled Virtue: An Inquiry into Moral Agency*. (New York, Oxford University Press, 2005)

Audi, R. *Actions, Intentions and Reasons*. (Ithaca, Cornell University Press, 1993)

Bentham, J. 'Punishment and Deterrence' from *Principled Sentencing* ed. Andrew von Hirsch and Andrew Ashworth (Edinburgh, Edinburgh University Press, 1992) p. 62-66

Braude, S. 'Multiple Personality and the Structure of the Self' from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p.134-144

Brock, D. 'Desert, Fairness and Persons' *Noûs*, Vol. 17, No. 1, 1983 A. P. A. Western Division Meetings (Mar., 1983), pp. 56-58

Buss, S. *Contours of Agency: Essays on themes from Harry Frankfurt* (Cambridge, MIT Press, 2002)

Butler, J. 'Of Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 99-106

California Press, 1976) p. 139-174

Campbell, C.A. 'Professor Smart on Free-Will, Praise and Blame; A Reply' *Mind*,
New Series, Vol. 72, No. 287 (Jul., 1963), pp. 400-405

Carithers, M. and S. Collins and S. Lukes (ed.) *The Category of the person: Anthropology, Philosophy, History* (New York, Cambridge University Press, 1985)

Chalmers, D. 'The Puzzle of Conscious Experience' from P. Van Inwagen and D.
Zimmerman (ed.) *Metaphysics: The Big Questions* (Malden, Blackwell Publishers, 1998) p. 333-342

Chappell, V. 'Locke and Relative Identity' *History of Philosophy Quarterly*, vol. 6,
pp. 69-83, January 1989

Chisolm, R. 'The First Person' from D. Rosenthal (ed.) *The Nature of Mind* (New
York, Oxford University Press, 1991) p. 312-325

Chisolm, R. 'Which Physical Thing Am I?' from P. Van Inwagen and D.
Zimmerman (ed.) *Metaphysics: The Big Questions* (Malden, Blackwell Publishers, 1998) p. 291-296

Chisolm, R. "Freedom and Action" in K. Lehrer (ed.) *Freedom and Determinism*
(New York, Random House, 1966)

Churchlad, P. *Matter and Consciousness* (Cambridge, MIT Press, 1988)

Crabtree, A. *Multiple Man: Explorations in Possession and Multiple Personality*
(Toronto, Praeger Publication, 1985)

Clarke, R. 'Toward A Credible Agent-Causal Account of Free Will' *Noûs*, Vol. 27,
No. 2 (Jun., 1993), pp. 191-203

Davidson, D. 'Agency' from *Essays on Actions and Events* (New York, Oxford University Press, 1980) p. 42-63

Davidson, D. 'Actions, Reasons and Causes' from *Essays on Actions and Events*

(New York, Oxford University Press, 1980) p. 3-20

De Sousa, R. 'Rational Homunculi' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 217-238

Dennett, D. 'Conditions of Personhood' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 175-196

Dennett, D. 'The Self as a Narrative Center of Gravity' from *Philosophical Horizons* (Belmont, Wadsworth Publishing, 2005) p. 125-131

Dennett, D. "I Could not have Done Otherwise -- So What?" Journal of Philosophy, vol. 8, pt. 1, pp. 553-565, October 1984

Dennett, D. 'Intentional Systems' *Journal of Philosophy*, vol. 68, pp. 87-106, February 1971

Descartes, R. 'Meditation II' from C. Landesman and R. Meeks (ed.) *Philosophical Skepticism* (Malden, Blackwell Publishing, 2003) p. 192-201

Diagnostic and Statistical Manual for Mental Disorders IV-TR Fourth Edition

Dworkin, G. "Acting Freely", Noûs, vol. 4, pp. 367-383, November 1970

Eshleman, A. 'Moral Responsibility' from *Stanford Encyclopedia of Philosphy http://plato.stanford.edu/entries/moral-responsibility/*

Eshleman, A. 'Being is not Believing: Fischer and Ravizza on Taking Responsibility' *Australasian Journal of Philosophy* Vol. 79 No. 4 p. 479-490; *December 2001*

Feldman, F. 'Desert: Reconsideration of Some Received Wisdom' *Mind*, New Series, Vol. 104, No. 413 (Jan., 1995), pp. 63-77

Fischer, J.M. 'Recent Work on Moral Responsibility' *Ethics*, Vol. 110, No. 1 (Oct., 1999), pp. 93-139

Fischer, J.M. 'Responsibility and Control' *Journal of Philosophy* Vol. 79, No. 1 (Jan. 1982), p 24-40

Fischer, J.M. and J. Ravizza *Responsibility and Control: A Theory of Moral Responsibility*. (New York, Cambridge University Press. 2000)

Fischer, J.M. *The Metaphysics of Free Will* (New York, Blackwell Publishing, 1996)

Flew, A. 'Locke and the Problem of Personal Identity' *The Journal of the Royal Institute of Philosophy*, vol. 2, pt. 6, pp. 53-68, January 1951

Frankfurt, H. 'Alternate Possibilities and Moral Responsibility' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p 1-10

Frankfurt, H. 'Coercion and Moral Responsibility' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p. 26-46

Frankfurt, H. 'Freedom of the Will and the Concept of a Person' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p. 11-25

Frankfurt, H. 'Identification and Externality' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p. 58-68

Frankfurt, H. 'Identification and Wholeheartedness' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p. 159-176

Frankfurt, H. 'Necessity and Desire' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p. 104-116

Frankfurt, H. 'The Importance of What We Care About' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p. 80-94

Frankfurt, H. 'The Problem of Action' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p. 69-79

Frankfurt, H. 'Three Concepts of Free Action' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p. 47-57

Frankfurt, H. 'What We Are Morally Responsible for' from *The Importance of What We Care About* (New York, Cambridge University Press, 1994) p.

95-103

Frankfurt, H. ‘Reply to Velleman’ from *Contours of Agency: Essays on Themes from Harry Frankfurt* edited by S. Buss and L. Overton (Cambridge, MIT Press, 2002) p. 124-128

Gordon, R. ‘Emotions and Knowledge’ from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 450-454

Grant, G. ‘Brain Bisection and Personal Identity’ *Mind: A Quarterly Review of Philosophy*, vol. 9, pt. 5, pp. 224-229, April 1986

Grice, H.P. ‘Personal Identity’ from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 73-98

Hilgard, E. ‘Dissociative Phenomena and the Hidden Observer’ from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p. 89-115

Honderich, T. ‘Compatibilism, Incompatibilism, and the Smart Aleck’ *Philosophy and Phenomenological Research*, Vol. 56, No. 4 (Dec., 1996), pp. 855-862

Hume, D. *A Treatise of Human Nature* by Ed E. Mossner (London, Penguin Books, reprint 1969)

Humphreys, N. and D. Dennett ‘Speaking for Ourselves: An Assessment of Multiple Personality Disorder’ from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p. 144-162

James, W. *Principles of Psychology Vol. 1* (New York ,Cosimo Classics, 2007)

Johnston, M. ‘Human Beings’ *Journal of Philosophy*, Vol. 84, No. 2 (Feb., 1987), pp. 59-83

Kant, I. 'Justice and Punishment' from *Philosophical Perspectives on Punishment* ed. Gertrude Ezorsky (New York, State University of New York Press, 1972)

Kant, I. ‘The Paralogisms of Pure Reason’ from C. Landesman and R. Meeks (ed.) *Philosophical Skepticism* (Malden, Blackwell Publishing, 2003) p.

214-219

Kant, Immanuel *Groundwork for the Metaphysics of Morals* Tr. Thomas E. Hill Jr.
and Arnulf Zweig (New York, Oxford University Press, 2002)

Kennett, J. and S. Matthews 'Identity, Control and Responsibility: the case of Dissociative Identity Disorder' by, *Philosophical Psychology,* vol. 15, no. 4
2002

Korsgaard, C. 'Personal Identity and the Unity of Agency: A Kantian Response to
Parfit' from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p. 323-338

Leibniz, G. *New Essays Concerning Human Understanding* Ed. J. Bennett and P.
Remnant (New York, Cambridge University Press, 1996)

Leibniz, Gottfried "Discourse on Metaphysics" Tr, R.S. Woolhouse and Richard
Francks, *G.W. Leibniz Philosophical Texts* (New York, Oxford University
Press, 1998)

Lewis, D. 'Postscripts' from D. Kolak and R. Martin (ed.) *Self and Identity* (New
York, Macmillan Publishing Company, 1991) p. 285-288

Lewis, D. 'Survival and Identity' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 17-31

Lewis, D. 'Counterparts of Persons and their Bodies' *Journal of Philosophy*, Vol.
68, No. 7 (Apr. 8, 1971), pp. 203-211

Lizza, J. 'Multiple Personality and Personal Identity Revisited' *The British Journal*
for the Philosophy of Science, Vol. 44, No. 2 (Jun., 1993), pp. 263-274

Locke, John *Essay Concerning Human Understanding* Ed P.N. Nidditch (Oxford,
Oxford University Press, 1979)

Loughery, D. 'Second-order Desire Accounts of Autonomy' *International*

Journal
of Philosophical Studies, vol, 6 (2), 211-229 1998

Ludwig, AM, Brandsma, JM, Wilbur, CB, Bendfeldt, F and DH Jameson. 'The Objective Study of a Multiple Personality' *Archives of General Psychiatry*
April 1972, 26 (4) 298-310

Macintyre, A. 'Determinism' *Mind*, New Series, Vol. 66, No. 261 (Jan., 1957), pp.
28-41

Malcolm, N. 'Thoughtless Brutes' from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 454-462

Martin R. 'Personal Identity and Survival: Becoming the persons we most want to
be'. In D. Kolak and R. Martin (eds.) *The Experience of Philosophy*. (New
York, Oxford University Press, reprint 2005)

Martin, R. 'Identity, Transformation, and What Matters in Survival' from D. Kolak
and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p. 289-300

Mauss, M. 'A Category of the Human Mind: The Notion of Person: The Notion of
Self' tr. H.D. Walls from Carithers, M. and S. Collins and S. Lukes (ed.) *The Category of the person* (New York, Cambridge University Press, 1985) p.1-25

Mele, A. 'Akrasia, Self-Control and Second Order Desires' *Noûs*, Vol. 26, No. 3
(Sep., 1992), pp. 281-302

Nagel, T. 'War and Massacre' *Philosophy and Public Affairs*, Vol. 1, No. 2 (Winter, 1972), pp. 123-144

Nagel, T. 'Brain Bisection and the Unity of Consciousness' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 227-
246

Nagel, T. *The View From Nowhere* (New York, Oxford University Press, 1986)

Neely, W. 'Freedom and Desire' *The Philosophical Review*, Vol. 83, No. 1 (Jan., 1974), pp. 32-54

Nietzsche, F. from *Beyond Good and Evil* from C. Landesman and R. Meeks (ed.) *Philosophical Skepticism* (Malden, Blackwell Publishing, 2003) p. 219-222

Nozick, R. 'Coercion' from *Socratic Puzzles*. (Cambridge, Harvard University Press, Reprint 1999) p. 15-44

Nozick, R. 'The Closest Continuer View' from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p. 212-226

Nozick, R. *Philosophical Explanations* (Cambridge, Harvard University Press, 1983)

Olson, E. 'Was Jekyll Hyde?' *Philosophy and Phenomenological Research* Vol. LXVI, No. 2, March 2003 p. 328-348

Olson, Eric, 'Personal Identity' from *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/entries/identity-personal/

Olson, E. *The Human Animal: Personal Identity Without Psychology* (New York, Oxford University Press, 1997)

Parfit, D. 'On 'The Importance of Self-Identity'' Journal of Philosophy, vol. 6, pt. 8, pp. 683-690, October 1971

Parfit, D. 'Divided Minds and the Nature of Persons' from P. Van Inwagen and D. Zimmerman (ed.) *Metaphysics: The Big Questions* (Malden, Blackwell Publishers, 1998) p. 310-317

Parfit, D. 'Lewis, Perry, and What Matters' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 91-109

Parfit, D. 'Personal Identity, Rationality, and Morality' from D. Kolak and R. Martin

(ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991)
p. 301 - 323

Parfit, D. 'Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 199-226

Parfit, D. 'The Psychological View' from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p. 227-267

Parfit, D. *Reasons and Persons*. (New York, Oxford University Press, 1984)

Penelhum, T. (1971) 'The Importance of Self-Identity' Journal of Philosophy, vol.
6, pt. 8, pp. 667-678, October 1971

Penelhum, T. 'Self-Identity and Self-Regard' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 253-280

Perry, J. 'Personal Identity, Memory and the Problem of Circularity' from J. Perry
(ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 135-158

Perry, J. 'The Importance of Being Identical' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 67-90

Perry, J. 'The Problem of Personal Identity' from J. Perry (ed.) *Personal Identity*
(Berkeley. University of California Press, 1975) p. 3-32

Petit, Philip and Michael Smith 'Freedom in Belief and Desire' *Journal of Philosophy* Vol. 3 No. 9 (Sep. 1996) p. 429-449

Prince, M. *The Dissociation of a Personality* (New York, Greenwood Press, 1969
Reprint)

Pucceti, R. 'Brain Bisection and Personal Identity' *British Journal for the Philosophy of Science*, vol. 2, pt. 4, pp. 339-355, December 1973

Quinton, A. 'The Soul' from J. Perry (ed.) *Personal Identity* (Berkeley. University
of California Press, 1975) p. 53-72

Rawls, J. *A Theory of Justice* (Cambridge, Belknap Press, 1972)

Reid, T. 'Of Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 107-112

Reid, T. 'Of Mr. Locke's Account of Our Personal Identity' from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 113 – 118

Rey, G. 'Survival' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 41-66

Rorty, A.O. 'A Literary Postscript: Characters, Persons, Selves, Individuals' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 301-324

Rorty, A.O. Introduction, from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 1-16

Rorty, R. 'Persons without Minds' from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 268-287

Rosenthal, D. 'Two Concepts of Consciousness' from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 462-479

Rosenthal, D. 'State Consciousness and Transitive Consciousness' *Consciousness and Cognition*, 2, 4 (December 1993), pp. 355-363

Rosenthal, D. *Consciousness and Mind (*Oxford: Clarendon Press, 2005)

Russell 'Strawson's way of naturalizing responsibility' *Ethics,* Vol. 102, No. 2 (Jan. 1992) pp. 287-302

Ryle G. 'Self Knowledge' from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 429-432

Schechtman, M. *The Constitutions of Selves*. (Ithaca, Cornell University Press, 1996)

Schechtman, M. 'Personhood and Personal Identity' *Journal of Philosophy*, Vol. 87, No.2 (Feb. 1990) p. 71-92

Scheffler, S. ‘Responsibility, Reactive Attitudes, and Liberalism in Philosophy and
Politics’ *Philosophy and Public Affairs*, Vol. 21, No. 4 (Autumn, 1992), pp.
299-323

Schlick, M. “When Is a Man Responsible?” in *Problems of Ethics* tr. David Rynn.
(New York, Prentice Hall, 1939) pp. 143-156

Searle, J. *Rationality in Action* (Cambridge, MIT Press, 2001)

Shaffer, J. ‘Personal Identity: The Implications of Brain Bisection and Brain Transplants’ *Journal of Medicine and Philosophy*, vol. 2, pp. 147-161, June 1977

Shoemaker, S. ‘Embodiment and Behavior’ from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 109-138

Shoemaker, S. ‘Personal Identity and Memory’ from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 119 – 134

Shoemaker, S. ‘Survival and the Importance of Identity’ from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p. 267-272

Shoemaker, S. and R. Swinburne *Personal Identity*. (Oxford, Basil Blackwell, 1984)

Smart, J.J.C. ‘Free-Will, Praise and Blame’ *Mind, New Series*, Vol. 70, No. 279,
291-306. Jul., 1961. VOL. LXX. No. 279

Smilansky, S. ‘Responsibility and Desert: Defending the Connection’ *Mind*, New
Series, Vol. 105, No. 417 (Jan., 1996), pp. 157-163

Stern , L. ‘Freedom, Blame, and Moral Community’ *Journal of Philosophy*, Vol.
71, No. 3, 72-84

Strawson P. ‘Freedom and Resentment’ from *Free Will: Oxford Readings in Philosophy* by Gary Watson (ed.) (New York, Oxford University Press, 2003) p. 72-93

Strawson, G. 'Against Narrativity' *Ratio* Volume 17 Issue 4 Page 428-452, December 2004

Strawson, G. 'The Self and the Sesmet' *Journal of Consciousness Studies*, Volume 6, Number 4, 1999 , pp. 99-135(37)

Strawson, P.F. 'Persons' from D. Rosenthal (ed.) *The Nature of Mind* (New York, Oxford University Press, 1991) p. 104-116

Stump, E. 'Sanctification, Hardening of the Heart and Frankfurt's Concept of Free Will' *Journal of Philosophy*, Vol. 85, No. 8 (Aug 1988) p. 395-420

Taylor, C. 'Responsibility for Self' from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of California Press, 1976) p. 281-300

Unger, P. 'The Physical View' from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company, 1991) p. 192-211

Unger, P. *Identity, Consciousness and Value* (New York, Oxford University Press, 1992)

Van den Beld, T. 'Responsibility and Personal Identity' from *Personal and Moral Identity* edited by A. Musschenga, W. van Haaften, B. Spiecker, and M. Slors (Boston, Kluwer Academic Publishers, 2002)

van Inwagen, P. "Ability and Responsibility" Philosophical Review, vol. 8, pt. 7, pp. 201-224, April 1978

Velleman, D. "What Happens When Someone Acts?" in *Mind*, New Series, Vol. 101, No. 403, 1992, p. 461-481

Velleman, D. 'Identification and Identity' from *Contours of Agency: Essays on Themes from Harry Frankfurt* edited by S. Buss and L. Overton (Cambridge, MIT Press, 2002) p. 91-123

Velleman, D. 'Self to Self' *The Philosophical Review* Vol. 105 No. 1 (Jan.1996), p. 39-76

Wallace, R.J. *Responsibility and the Moral Sentiments* (Cambridge, Harvard

University Press, 1994)

Watson, G. ‘Free Agency’ Journal of Philosophy, vol. 7, pt. 2, pp. 205-220, April
1975

Wedeking, G. ‘Locke’s Metaphysics of Personal Identity’ *History of Philosophy*
Quarterly, vol. 4, pp. 17-31, January 1987

Wiggins, D. ‘Locke, Butler and Stream of Consciousness: And Men as Natural Kinds’ from A.O. Rorty (ed) *Identities of Persons* (Berkeley, University of
California Press, 1976) p. 135-174

Wilkes, K. ‘Fugues, Hypnosis and Multiple Personality’ from D. Kolak and R. Martin (ed.) *Self and Identity* (New York, Macmillan Publishing Company,1991) p.115-134

Williams, B. ‘Persons, Character and Morality’ from A.O. Rorty (ed) *Identities of*
Persons (Berkeley, University of California Press, 1976) p. 197-216

Williams, B. ‘The Self and the Future’ from J. Perry (ed.) *Personal Identity* (Berkeley. University of California Press, 1975) p. 179-198

Wittgenstein, L. from *Tractatus Logico-philosophicus* from C. Landesman and R.
Meeks (ed.) *Philosophical Skepticism* (Malden, Blackwell Publishing, 2003) p. 222-225

Wolf, S. ‘The Importance of Free Will’ from *Perspectives on Moral Responsibility*
ed. By J.M. Fischer and Mark Ravizza (New York, Cornell University Press, 1993) p. 101-118