Department of Philosophy

University of the Western Cape

UNIVERSITY *of the*
WESTERN CAPE

Master of Arts (coursework) Research Project

A mini thesis submitted in partial fulfilment of the requirements for the degree of Master of Arts, submitted to the Department of Philosophy, University of the Western Cape.

Personal identity, consciousness and a self-concept.

Candice Johnstone
3431588

Supervisor: Professor Simon Beck

I, Candice Johnstone, hereby declare that *Personal identity, consciousness and a self-concept* is my own original work and that all sources have been accurately reported and acknowledged, and that this document has not previously in its entirety or in part been submitted at any university in order to obtain an academic qualification.

Personal identity, consciousness and a self-concept

Abstract

With a universally agreed upon perspective towards personal identity yet to be discovered, philosophers continue investigating the metaphysical question of what it means for a person to be one and the same person over time. Current views include those offered by those in the Lockean tradition, who claim that who we are lies within memories (Locke, 1975), while Parfit (1984) furthers this view to require psychological continuity, or what he calls *Relation R.* In contrast, contributors such as Williams (1970) and Olson (1997a) postulate that bodily continuity is responsible for the sameness of self. Schechtman (1996) puts forward a more contemporary perspective emphasising a narrative view, particularly as a component of her combination theory. The aim of this paper is to evaluate the work of each of the above-mentioned theorists and expose the damaging charges their respective views of personal identity are confronted with. In the hopes of addressing the theoretical gaps we are therefore left with, I will propose my own view of where our personal identity lies, borrowing essential concepts from Locke's use of consciousness and Schechtman's narrative view and self-concept, highlighting that a person is required to be aware of their existence within the world. This perspective will work to avoid the charges against the central theories identified above, while offering insight into the practical and ethical implications of a view of personal identity relying on consciousness and a self-concept.

Key words

Personal identity, personhood, memory, bodily continuity/criterion, psychological continuity/criterion.

Introduction

The philosophical concept of personal identity has had philosophers occupied and intrigued for centuries. Philosophers have been working to pinpoint what it is that makes a person the same person over time. With a variety of contrasting perspectives currently on offer, we are yet to come to a conclusive agreement of what it is that makes a person the same person over time. Aside from a theoretical standpoint, the question of personal identity carries an ever-increasing practical relevance. From hypothetical thought experiments to potential real-world head transplants, the question of who survives is becoming a more legitimate and real concern.

The discussion to follow will investigate several views that philosophers over time have developed in an attempt to identify the criterion for personal identity. Examining every individual theory of personal identity will prove to be excessive and superfluous. For this reason, and considering time constraints, only the primary and most prominent theories of personal identity, will be explored in this research project. Included in the investigation of several views of persisting personal identity, it will be demonstrated by using thought experiments and identifying respective challenges why each theory discussed below offers a criterion for personal identity that is implausible and unconvincing and should therefore be rejected as the criterion for persisting personal identity. The strategy to illuminate the concerns and ultimately the path to rejecting each of the below theories is to demonstrate how each theory, individually, fails to do what a plausible theory of personal identity is supposed to accomplish. The requirements of a plausible theory of personal identity will be detailed before investigating the theories themselves, although the four features will not be examined against the views offered by Marya Schechtman that will be considered later in the discussion.

Before delving into the various approaches on offer to explain personal identity, it is vital to articulate what exactly is meant by concepts such as personal identity, and to acknowledge that while philosophers often speak of personal identity and survival, these two concepts are not

synonyms. It is important to clarify that the concept of personal identity will be explored through a metaphysical philosophical lens, and that the concept of personal identity is not the same notion as *personality* the way we understand in the field of psychology. As will be discussed in more detail later, personal identity, as argued by Marya Schechtman (1996; 2014), should be seen as a concept where two different elements run together, namely the self and the person, where the self refers to the personal or survival of oneself, and the concept of *person* applies to being the same person to *other people* (the same person as I am from others' perspectives). Survival, on the other hand, is a matter of an individual persisting over time. A closer look into these finer concepts will be offered as to distinguish their respective uses and relationships before examining the theories that are to be included in the discussion to follow. Finally, the four features that a theory of personal identity needs to address, namely survival, moral responsibility, self-interest concern, and compensation, will be defined and incorporated into the conversation about what a criterion of personal identity should accomplish, before analysing how well the theories in fact capture these four features.

In the late 1600s, John Locke offered the first systematic study of the subject of personal identity in setting out his view that our consciousness is responsible for the sameness of a person "since consciousness always accompanies thinking, and '*tis that, that makes every one to be, what he calls self; and thereby distinguishes himself from all other thinking things, in this alone consists personal Identity, i.e., the sameness of a rational Being*'" (Locke, 1975: 335). Consciousness here is not synonymous with memory and Locke is not requiring an individual to possess memories of each particular moment and event in one's life. A sufficient stream of consciousness is achieved when a person possesses memories of themselves from previous moments in their life where each memory is connected to said moment. Locke (1975) requires that either we can recall a previous experience, or that we can be made to do so. Instead, sameness of consciousness refers to the continuation of a person's consciousness provided most obviously by chains of memories. Remnants of Locke's original approach can be located in several more contemporary theories of personal identity. Derek Parfit (1984) uses Locke's memory criterion as a foundation to develop one of the most well-known

and recognized views on personal persistence where he considers psychological continuity as the necessary criterion and puts forward the concept of *Relation R* as what matters in personal survival. Parfit's *Relation R* can be described as psychological continuity and connectedness which refers to "the holding of direct psychological connections" (1984: 206). Instead of requiring direct memories or being made to recall direct memories as witnessed with Locke, Parfit requires a type of psychological continuity that is formed with "overlapping chains of *strong* (psychological) connectedness" and explains that there "*is enough connectedness if the number of direct connections, over any day, is at least half the number that hold, over every day, in the lives of nearly every actual person"* (1984: 206).

Jeff McMahan argues for an approach where the criterion is neither strictly physical nor psychological. McMahan's Embodied Mind View requires a degree of physical continuity that supports "the continuity of the capacity for consciousness" (Oyowe, 2010: 275). For McMahan, the *same brain* is required to support the *functional continuity* that supports the *capacity* for one's consciousness (McMahan; 2002). Important to note is that McMahan does not require continuity or possession of one's mental content (Oyowe, 2010: 275). Rather, it is the *capacity* for consciousness provided by the same physical brain that McMahan asserts is necessary for continued personal identity.

Contrary to these psychological views (as well as McMahan's Embodied Mind Approach), English philosopher, Bernard Williams (1970), offers an approach which significantly opposes the suggestion of personal identity being explained by criteria such as an individual's memories or psychology. Williams considers bodily continuity as the required condition for personal identity. Importantly, Williams is not arguing for the continued existence of a person's entire physical body. Instead, Williams's view requires continuity of *enough* of the body that is able to support psychological and mental functions and capabilities. Additionally, Williams does not, unlike many psychological theorists, depend on the continuity or even presence of particular mental or psychological content.

Williams's view, importantly, differs from a psychological approach where he uses a familiar thought experiment to support his view that a person can and does continue to exist as the same person albeit experiencing tremendous changes to their psychology. Animalist, Eric Olson (1997a), suggests a criterion that shares some level of resemblance with Williams. Olson argues for a biological criterion; however, he does not require continuity of one's (entire) physical body. Rather, Olson rejects any suggestion of a non-physical or psychological criterion, requiring instead that your persistence is a matter of *organism* continuity. Although he also refers to his view as a theory of personal identity, Olson's organism view, or biological approach, requires the continued existence and functioning of one's brainstem. Olson prioritizes the brainstem as it is this structure that organises one's body and thus argues that it makes you the same organism. Olson (1997a) argues that the necessary and sufficient criterion for persisting personal identity is to possess the same functioning brainstem as it facilitates and maintains the basic human processes and operations such as breathing and heart rate regulation. Provided a person possesses their original and functioning brainstem, they satisfy the requirement for personal persistence.

We will then move on to two contributions proposed by Marya Schechtman who offers an earlier view (1996) and a later view (2014), both of which vary greatly from the abovementioned approaches. Although her latest view (2010; 2014) includes both physical and psychological continuities, Schechtman argues that these alone are not sufficient for personal identity. Schechtman's (1996) earlier (Narrative Self-Constitution View, or the Self-Understanding) view considers personal identity to be a matter of a person's capacity to construct a coherent and consistent story, or narrative, of themselves and their life. According to her narrative view, Schechtman argues that "to be the same person is to have a particular self-understanding or a 'sense of self' which involves seeing and understanding experiences and actions as part of an intelligible whole" (Schechtman, 1996). The narrative view explains that an experience, thought, or action is *yours* only if you can logically fit it into the narrative of your life in a meaningful manner. Where Schechtman differs from her counterparts is the fact that her view appears much more malleable.

What is meant by this is that a person does not have to "fit" into the criterion of bodily or psychological continuity. Schechtman (1996) thinks the above theories focus on the wrong issue – that of re-identification rather than characterization (the latter being more closely related to the 'four features' that are the concern of personal identity). Instead, Schechtman's narrative view seems to allow personal identity to remain despite some inevitable (physical and psychological) changes and developments a person experiences throughout their lifetime, a flexibility which the above physical and psychological continuity theories (PCT) fail to permit. Additionally, Schechtman accuses bodily and psychological continuity theorists of omitting a fundamental element of personal identity. That is, the social element namely the ability of a person to engage with other people, is a crucial element to personal identity, according to Schechtman. Her Person Life View, defined as being "a person is to live a 'person life'; persons are individuated by individuating person lives; and the duration of a single person is determined by the duration of a single person life" (2010; 2014: 110), demonstrates a shift from an emphasis on a person's own narrative to the narrative of those around them (society) in what she refers to as a person's 'place in person-space'. Each person lives their own individuating person lives and interact with those around them in such a way that society recognizes you as the same person over time (instead of you yourself understanding and recognizing that you are the same person over time) (Beck, 2016: 128). These two views from Schechtman will be expanded on and assessed against the four features. Nearing the end of this paper is when we will examine the potential resolution to the debate of personal identity where we will consider Schechtman's narrative view as working better than her PLV.

Purpose of research and relevance of research question

As mentioned in the beginning, there is more to the conversation around personal identity than simply theoretical arguments speculating as to what makes a person one and the same over time. With the advancement of modern technology, we are edging closer to the possible reality of

scenarios we have thus far only considered to be hypothetical thought experiments. Considering the possibility of the presence of potential ethical concerns being involved, the question of personal identity starts requiring an answer for practical reasons. In an unprecedented procedure, Italian neuroscientist, Dr Sergio Canavero, announced in 2013 that he and Chinese Surgeon, Dr Xiaoping Ren, would have performed the world's first head transplant where Russian volunteer, Valery Spiridonov would have his head separated from his body and transplanted onto a headless donor body, in an attempt to allow him to live a normal, healthy life after suffering from Werdnig-Hoffmann Disease, a rare, incurable, and often fatal genetic disorder that leads to the degeneration and death of nerve cells of the brain and spinal cord (Melnick, 2017). A procedure like this not only raises concerns around ethics and morality, but depending on the theory one subscribes to, the question as to who will in fact survive and wake up from the surgery yields opposing answers. Logically, if an operation like this were to take place, we would *need* an answer as to who survives, for practical and legal, amongst other reasons.

Going beyond a meta-analysis of some of the most well-recognised approaches to personal identity and examining where they fail, this research project aims to put forward an alternative set of criteria necessary for personal persistence. I intend to achieve this by borrowing a concept from Locke, that being our consciousness, as it can be argued this element has an essential role in one's continued identity. However, as I will demonstrate, consciousness, or continuity of one's consciousness alone is not sufficient for personal identity. Instead, I will use the concept of consciousness as a working criterion, where the factor of consciousness as Locke describes will not simply be replicated and advocated for. Rather, my alternative view will build on the fundamental basics of Locke's approach by suggesting that while (the stream of) consciousness is one of the necessary conditions for personal identity, so is a continued sense of self, or a self-concept, a notion which I will borrow from Schechtman. Where my alternative view differs from Locke is by dismissing the requirement of memories, both direct memories (memories to a specific experience) and the conception of ancestral selves (the causal relation between one's earlier and later consciousness) to

connect experiences. The purpose of rejecting the need for ancestral selves is to avoid the frequent criticism Locke faces when confronted with the inescapable event of forgetfulness, false memories, and significant gaps in memory (and the chain of memories) such as suffering from dementia, as well as not having a memory or a connection leading back to having been born.

My alternative view will work to reject other theories that rest in the foundation of a purely physical, bodily criterion. Where Olson considers the requirements for personal identity being satisfied if the person possesses their intact and functional brainstem, I reject this as insufficient for practical reasons, namely where a person in a vegetative state (therefore breathing and having other basic bodily functioning) can in fact open their eyes but demonstrate no sense of personal survival. This penetrative lack of self-awareness leads to, for many of us, the intuitive belief that although we rationally acknowledge the continued functioning of the organismic body, the *person* we once knew is no longer living. Furthermore, this alternative view will work to avoid the challenges Schechtman's social criterion faces by putting less emphasis on her external social aspect of personal identity, that way limiting the influence of the social criterion, but still acknowledging its importance. In comparison, the alternative view will include the proposal of a second criterion (the first being continued consciousness) of an internal sense of self where a person recognizes and acknowledges themselves as an existing person as well as a person who fits into society one way or another. Therefore, for persisting personal identity, the necessary and sufficient criteria include continuity of one's consciousness as well as being consciously aware of their own existence and place in society and reality, irrespective of possessing any other psychological contents, such as memories. This way, we are capable of ascribing features to a person with surety that we are ascribing these features to the *correct* person, that is if the feature fits coherently and meaningfully into the person's narrative (story) of their life. Here we are able to explain *what* makes a person the same person today as they were yesterday, while avoiding the need reidentification theorists face of explaining *how* a person the same person at t_1 one and the same person at t_2.

The four features of Personal Identity

When working with the concept of personal identity, we are met with a set of intuitions about why the concept of personal identity is such an important one. Marya Schechtman (1996: 2) argues that traditional theories of personal identity – those attempting to explain what makes a person at t_1 one and the same person at t_2 – are of the belief and understanding that "facts about identity underlie facts about four basic features of personal existence: survival, moral responsibility, self-interested concern, and compensation." Those working to provide an answer to the question of personal identify, or reidentification theorists, are under the assumption that "an acceptable criterion of personal identity should be able to capture *all* of our most basic intuitions" (Schechtman, 1996: 13). Schechtman (1996: 2) explains that the assumption that an explanation of personal identity needs to cover all the four features arises because each of these four features are linked to identity. Reidentification theorists assess the credibility and plausibility of a proposed criterion of personal identity based on the connection it makes between the four features and personal identity (1996: 2). It is important to note that Schechtman finds this assumption of the reidentification theorist deeply troubling, a concern that will be returned to later. For now, let us closely examine those four features of personal identity.

Arguably, the most important feature is survival and its connection to personal identity (Schechtman, 1996: 14). The feature of survival is concerned with whether or not a person will still exist in the future and is a matter of an individual persisting over time (1996). As the name of this feature suggests, here we are concerned with whether a person today will survive (continue to exist) in the future.

Next, facts about personal identity play a critical role in assigning moral responsibility (Schechtman, 1996: 14). In scenarios where responsibility needs to be assigned, such as identifying the culprit of a recent robbery, we need to know *who* it was who committed this crime and it is *that person* who subsequently needs to be held responsible for the misdeeds that *they* committed. We

place importance of knowing *who* it is that committed an action as it is only fair that the person being held responsible did in fact perform the action, and that this person is likewise (and appropriately) held responsible for only those prior actions *they* committed.

It is incontestable that as persons, we feel a special kind of rational interest (self-interested concern) about our own future (Schechtman, 1996). Schechtman (1996) explains that the concern we have for ourselves and our own future is of a special and particular kind because although we care deeply about others, and even just as much, there is a type of caring for oneself and one's future that rationally pertains only to themselves and differs from the care we feel for those around us. Schechtman differentiates the care we have for ourselves from the care for others by using the concept of anticipation where she explains that unlike those of others, we "expect to *feel* our own experiences" (1996: 14) as opposed to being unable to feel someone else's (expected) experience. Consider the following scenario: Both myself and a friend are having our wisdom teeth surgically removed the following week. While I care deeply about my friend and wish them a successful surgery and a recovery that is not too painful, the concern I feel about my own surgery is undeniably different. Although I anticipate my friend to be in pain immediately after the surgery, there is a different kind of concern that I feel in anticipation of the pain I will experience after my own respective surgery. This different *type* of concern is what self-interested concern is referring to.

Our discussion about self-interested concern flows neatly into the fourth and final feature of identity – compensation. Schechtman (1996: 14) explains that because we feel a type of self-interest for ourselves and for our own future, it only makes sense that we are concerned with compensatory fairness. This feature is characterised by compensating a person based on the sacrifices *they* made (1996). However, to compensate a person – the *correct and deserving* person – we need to know who the person is that made the sacrifice and therefore earned the compensation. A clear example would be the ordinary, everyday scenario of salaries being paid. When I perform duties at my workplace, it is only fair that *I* am the person who receives the compensatory salary for the sacrifices

I made of giving up my time and energy to be at work. Of course, it is the case that *I* should receive the salary, and not anyone else (unless they too made the sacrifices and would therefore earn their own respective salary, not *my* earned salary).

With the four features thoroughly detailed above, we will now move on to examining a set of differing approaches to personal identity and assess how well they each respectively capture these four features of identity. Thereafter, Schechtman's concerns with the connection between personal identity and these four features will be investigated.

Persons and Consciousness

Before we examine several views debating where personal identity lies and being able to determine whether a person persists as *the same person* over time, let us first consider what is meant by *person* and *consciousness.* John Locke (1975: 335) describes a *person* as "a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places". I am a *person* as I am an existing being with the capacity to think and reason at an intellectual level, and to reflect on my existence as being my own existence, the same *me* today as I was yesterday. In his book, *The Ethics of Killing,* Jeff McMahan (2002: 45) describes a *person* as "a being with a rich and complex mental life, a mental life of a high order or sophistication". Similarly, Derek Parfit (1984: 202) theorizes that "to be a person, a being must be self-conscious, aware of its identity and its continued existence over time".

It is worth noting that of these three conceptions of a *person,* all have suggested that there is more to being a *person* than merely existing as the physical organism we regard as a human being. The above three notions of a *person* appear to include a connection to a concept of *consciousness.* Shelley Weinberg (2011) complains that it is not entirely clear what Locke referred to when speaking about one's *consciousness.* Weinberg (2011) questions whether Locke is possibly referring to (1),

memories, (2) a first person appropriation of mental states, or (3) a first personal distinguishing experience of the (qualitative) elements of one's own thinking. Locke does, however, suggest that having the capability to think, reason, and reflect stems from one's consciousness, and considers consciousness as "inseparable from thinking" and essential to thinking (1975: 335). Locke (1975: 335) puts this neatly as it is "impossible for any one to perceive, without perceiving, that he does perceive." It is clear that in Locke's description, there is a clear distinction between thinking and consciousness, although Locke considers the two closely connected (Weinberg, 2011). It follows that within Locke's conceptualisation of *consciousness,* the idea of self-consciousness is included, as Locke is requiring that *persons* are aware that *they* are the ones perceiving (Weinberg, 2011).

McMahan is in agreement with Locke by suggesting that a being that lacks self-consciousness "could not have a mental life that was strongly psychologically connected from day to day" (2002: 45). Again, we witness a need of a consciousness in the conception of a *person.* McMahan (2002: 45) explains that self-consciousness is needed for mental content and functions such as our beliefs, memories, desires, future intentions, and so on, and that these are the elements necessary for psychological connectedness. Interestingly, if one's consciousness, particularly their *self-consciousness* is required to be a *person,* it then follows that those (human) beings who do not yet, or no longer, such as infants and Alzheimer's patients respectively, possess the capacity for self-conscious thought are not considered *persons* (McMahan, 2002: 25). Moreover, as our psychological development is gradual, it is unknown when a being's ability for self-conscious thinking formally exists (McMahan, 2002: 45). McMahan (2002: 46) uses this to consider the notion of *person* as merely a *phase* in the history of a human being. Having introduced the concepts of *persons* and *consciousness,* a look into various theories of personal identity will follow, beginning with the work of John Locke.

Psychological theories of Personal Identity

John Locke and Memory

A name familiar to many of those working with the concept of personal identity is that of 17th century philosopher, John Locke. Locke (1975, Book 2, Chapter 27) offers a non-physical view to personal identity, defending his psychological approach to personal identity. Denying the importance of a physical criterion such as continuity of one's physical body, as well as the notion that personal identity can be a matter of having the same soul, Locke affirms that the essence of personal identity, or for a person to remain one and the same person over time, is the continuity of their consciousness, particularly of their memories. There are, however, two different questions where Locke says that we should first have clarity on what a person is (the nominal essence of identity as discussed above), then we can work on determining what the identity conditions of that kind (a *person*) will be. An illustration of Locke's psychological view is provided in the below description of the Prince and the Cobbler scenario, found in the 1975 version of his original *An Essay Concerning Human Understanding* (1694):

> *For should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, every one sees he would be the same person with the prince, accountable only for the prince's actions: but who would say it was the same man? The body too goes to the making the man, and would, I guess, to everybody determine the man in this case, wherein the soul, with all its princely thoughts about it, would not make another man: but he would be the same cobbler to every one besides himself.*

As thought experiments are employed to reveal where our intuitions lie with regards to personal identity, Locke (1975) contends that the above scenario demonstrates the importance and necessary psychological element of the continuity of our consciousness and memories to maintain personal identity. Locke (1975) not only sees the cobbler body as no longer housing the cobbler but instead the prince, but insists that it is, or at least should be *obvious* that the cobbler body is not the same

person as the cobbler, as the cobbler body now possesses a consciousness that is vastly different from the original cobbler person. In fact, the consciousness associated with the cobbler body is exactly identical to what we witnessed from the prince prior to the replacement of the cobbler's consciousness by that of the prince. While Locke says little about how the individual behaves, he thinks it is immediately obvious that once the (soul and) consciousness is transferred into the cobbler body, the person is the prince. The cobbler body now behaves and thinks in a way that is exactly similar to how the original prince thought and behaved, leaving Locke to argue that it only follows we see the new cobbler body person as the prince. Furthermore, while Locke (1975) acknowledges that the people around the cobbler might well think he is still the cobbler, we know what has taken place (the consciousness of the prince entering the cobbler's body) and thus judge that he (the cobbler body) is no longer the cobbler, and that the cobbler body person should not be held accountable for any actions, thoughts, or wrongdoings committed by the original cobbler as it was *not the same person* who actioned these deeds. Rather, because the *original* cobbler person no longer exists, the person in the cobbler body ought to be held responsible for the prince's actions, as it was *his* consciousness (that of the prince) that was occupying the cobbler's body at the time the actions were committed, not the consciousness of the cobbler. Moreover, any actions or misdeeds committed by the now cobbler body should be considered to be actions of the *original prince,* as this is where the prince's consciousness now lies. As mentioned, Locke (1975) grounds these claims in our intuitive reaction to the prince and the cobbler scenario, arguing that this thought experiment shows that continuity of the body is not essential to personal identity.

In anticipation of potential critique, Locke (1975) offers several refinements to the prince and the cobbler scenario. First, Locke (1975) recognises that our consciousness is not a constant, uninterrupted stream throughout the entirety of our lives, and permits natural breaks or gaps in one's consciousness. Locke (1975) admits that it is in fact not feasible for one's consciousness to never experience some form of break in continuity. The example of going to sleep and waking up illuminates the permissibility of temporary gaps in one's stream of consciousness where we naturally

and undoubtedly see someone as the same person when they wake up as the person who had gone to sleep, albeit that they would have experienced a momentary break in the continuity of their consciousness. In his theory, Locke (1975) emphasises the importance of memories and asserts that the person who wakes up is the same person as they possess the same memories as the person who went to sleep shortly before. In other words, a person continues to be the same person over time only if they possess the memories of themselves from previous moments in their life, and that each memory is connected to the actual experience that it is a memory of (Piccirillo, 2010). Locke (1975) accepts that we are not continuously conscious of all of our memories – we do not always recall them all and may actively remember none. It is enough that we *can* recall them or can be made to do so. Locke considers this an adequate defence against the charge of gaps in one's stream of consciousness, explaining that although a person might have temporary interruptions of consciousness, if they remember themselves committing an act, then it was indeed them who committed the act, despite the gap in consciousness. Using again the example of going to sleep, I am the same person waking up as the person who went to sleep because I remember having gone to sleep. Locke views this chain of memory links as sufficient to support the continuity of consciousness, and therefore maintaining personal identity.

Unfortunately for Locke, the specifics of his memory theory are not always entirely apparent. If what is necessary for continued personal identity is remembering all of one's actions from the past, Locke would run into trouble, according to Thomas Reid (1785, as cited in Locke, 1975: 114). Reid uses the below thought experiment, *The Brave Officer,* to demonstrate how the need to possess the same memories falls short in terms of the logic of transitivity (1785, as cited in Locke, 1975: 114):

> *Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have made a general in advanced life: Suppose also, which must be admitted to be possible, that when he took the standard, he was conscious of his having been flogged at school,*

> *and that when he made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.*

When considering the above example, the general, according to the interpretation of Locke's view as a simple memory theory, would not be the same person as the young boy (as he, the general, does not possess memories from the time he was a young boy who robbed the orchard). However, intuition leads us to believe that it is obviously the case that the general and the young boy *are* the same person, and that the general has merely forgotten the memories of being the young boy who robbed an orchard. This intuitive response presents a complication for Locke if his theory for being the same person were to solely require maintaining the same memories of our actions over time.

Let us examine how problems such as forgetting memories and the shortfall of the logic of transitivity can be addressed. Instead of interpreting Locke's view as a simple memory theory, we should consider Locke's understanding of what is meant by 'the same consciousness". By now, it is clear that the sameness of one's consciousness is not a result of having the same memories of one's actions throughout their lives, although you would be forgiven for interpreting Locke in this way. As Locke accurately asserts, we are not continually conscious of all our past actions; however, this does not lead to a person no longer having the same consciousness. This would mean that there need to be causal relations (other than memory) between earlier consciousness and later consciousness that also contribute to this being the same consciousness. Weinberg (2011) provides a neat explanation of a causal relation between earlier and later consciousness whereby (the sameness of) identity of a self that is otherwise forgotten can be established through an intermediate self (an ancestor self) that we do remember. Looking again at Reid's scenario, the brave officer, who remembers stealing from the orchard and who is also remembered by the general, would serve as the ancestral self, or the link, between the general today and the young boy from many decades ago, even though it is the case that the general today has no memory of robbing the orchard (Weinberg, 2011). The notion of an 'ancestral self' as a causal relation between earlier and later consciousness serves as potential

remedy against charges of forgetting memories and failing the logic of transitivity that Locke's view encounters.

A challenge that does not survive Locke's refined prince and Cobbler scenario is the fact that Locke's ancestral self (the causal relation between one's earlier and later consciousness) does not trace all the way back to when a person was born. Simply put, we have no memory of being born, either a direct memory or an ancestral self possessing a direct memory of an earlier self that remembers being born. By not having the memory (or connection to an earlier self who possesses the memory) of being born, we lack a connection to this event in our lives and therefore cannot claim this experience (of being born) as *our* experience. However, it would be nonsensical to say that *I* did not exist when I was born. Locke's distinction between the ideas of human being and person is precisely meant to solve this – it depends on what you mean by 'I' (Locke, 1975: 342). The relevant meaning when it comes to allocating responsibility is "I as a person". And that does not apply to you when you were born, according to Locke. Furthermore, the rule of identity dictates that two entities cannot have the same beginning, and one thing cannot have two beginnings, meaning that a person today cannot be a completely different person from who was born at the beginning of the life (of the physical body). This discrepancy reveals a troubling concern for Locke's memory theory, illustrating that his psychological criterion is not an adequate explanation of sameness of self. It should be noted that Locke's (1975) point here is that it was a human that was born, not a person, and that these are different kinds of things, with different identity conditions. There are, however, issues here – exactly how many entities are there once someone gains self-consciousness and becomes a person? When looking at a person who is self-conscious, are we in fact looking at two entities simultaneously? It appears somewhat counterintuitive for this to be the case. While this point will not be discussed further here, it highlights an interesting concern with Locke's psychological criterion.

In an attempted defence, Locke suggests that a newborn cannot be regarded as a person as at this stage, they are not yet self-conscious, proposing that human beings and persons are different kinds of things. This way, Locke is successful in avoiding the charge of two of the *same* things occupying the *same* place at the *same* time. Additionally, this defence follows what we naturally see in society today in that we do not hold those without a sense of self-awareness accountable for actions committed in this state. A simple example being that we do not hold a baby accountable for breaking their mother's precious glass vase in the same way we would hold a teenager accountable for this action. This suggests that Locke argues for a connection between personal identity and moral responsibility. However, if it is the case where human beings and persons are different kinds of things, a concern arises where there seem to be too many entities involved. A human being seems to be just one thing, or perhaps a combination of two things (soul and body). Yet, Locke appears to be making it three things, without providing an account of the metaphysical status of that third thing, other than to say personal identity is not a matter of either soul identity or body identity. This is certainly a little puzzling and raises a question about Locke's theory. More details of the issue here will be provided in the discussion of Olson's theory.

Locke's memory theory faces practical concerns in addition to the charges discussed above. Suppose an elderly woman suffers from a neurodegenerative condition such as Alzheimer's disease and experiences dementia where she has lost majority of her memories. According to Locke and his memory theory, the elderly woman is no longer the same person as she has no connection to the person who experienced the events of her life. It is not clear that the experiences are *hers.* However, this strongly contradicts what our intuition tells us about the personal identity of the elderly woman. Although the elderly woman has lost her memories, it is somewhat natural for those around her to still consider the elderly woman as existing. She would still be the same mother to her children, grandmother, and so on, albeit having lost her memories. Intuition tells us that we would not view the elderly woman as having ceased to exist. Suppose the elderly woman only experiences loss of a substantial portion of her memories from the last 20 years of her life, however older memories from

earlier in her life currently remain intact. Locke's memory theory is still not immune to the same charge. Locke's memory criterion would dictate that the elderly woman no longer exists due to the lack of a chain of memories linking the person today to the person who is connected to the experiences all those decades prior. Again, intuition plays a key role in demonstrating that we would in fact consider the elderly woman we see today (notwithstanding the partial memory loss) to be one and the same as the younger woman who experienced the events decades before. Furthermore, it would be nonsensical and illogical to suggest that either a different person, or no person at all, exists now in the place of the elderly woman, which is what Locke's memory criterion is forced to commit to. This scenario illustrates how Locke's view faces significant challenges and ultimately struggles to provide an explanation as to why we intuitively consider the elderly woman, albeit suffering from tremendous memory loss, to be the same person who continues to exist (Schechtman, 2010).

Before we close this section, let us assess the connection between Locke's approach and the four features of identity to ascertain whether his view is successful in accomplishing what a theory of personal identity ought to do. For the sake of time, the focus will be on those features that spark trouble for Locke. In terms of moral responsibility, recall this feature details that the appropriate person must be held responsible for actions *they* commit. Locke argues that a person cannot be held responsible for an action they have no memory of, and that it would in fact be wrong to punish anyone for an act that they cannot actively remember (Locke 1975: 346-7). This means that according to Locke, if a person cannot remember committing an action, they would therefore not be the right person to hold accountable. However, this contradicts our intuition as well as what we see in society. Suppose I rob a bank and although I have absolutely no recollection of this, there is clear video evidence. Regardless of possessing memory of this misdeed, I am held accountable in the eyes of society and the law, and this is the natural order of what we would expect to see happen. This highlights an incongruence between this feature and Locke's account, where Locke's view is unable to satisfactorily explain why we are happy to hold persons accountable irrespective of whether they

can remember committing the act, and ultimately fails to achieve this feature of identity. Interestingly, Locke (1975) envisages God restoring memories like these (robbing a bank) on the Day of Judgement which seems to place emphasis on the need of the ability to recall memory of an action in order to be fairly judged and held responsible for the action. Locke (1975) also recognises that human justice may be lacking because of our inability to access the actual consciousness of others (1975: 343-4). However, access to others' consciousness should not be what justice hinges on. Suppose we can in fact access one another's consciousness, and I discover my friend has absolutely no recollection of stealing my purse, however I saw her committing the crime when she thought I was not in the vicinity. I do not need access to my friend's consciousness to know that it was her who stole my purse and who should therefore be held responsible. Whether a person can recall performing an action or whether we can access one another's consciousness does not change how we intuitively feel about holding people accountable for the actions they undeniably committed.

The feature of compensation presents a similar challenge for Locke. As we have seen, Locke argues that only if a person remembers committing an action should they be punished, and that it would be wrong to hold a person accountable for actions they did not commit, and that God would not punish them for actions that are not their own. By virtue of this, a person should only be rewarded for a sacrifice they have the memory of making. Without remembering that it was *them* who made the sacrifice, Locke's account argues that we cannot tie the sacrifice to the person. This again would work against our intuition. For instance, imagine a friend of yours was in need of some money and without a moment of hesitation, I made the sacrifice of giving all the money I had on me to help my friend. Let us also imagine that when this happened, I had many other things going on and my attention was elsewhere which led me to forgot about giving my friend the money immediately afterwards. A week later, my friend calls and would like to take me out for an expensive dinner to thank me for helping them financially the previous week. As life happens sometimes, I have been so busy that I honestly have no recollection of giving my friend money. However, my

friend insists on taking me out to thank me. It seems misguided to say that *I* did not give my friend money last week just because I cannot remember doing so. I would not consider that person who gave the money a different person to who I am today, a week later. I may instead just be surprised at how I was able to forget this action I performed just a few days prior, but I would still see myself as the same, although more forgetful, person. Locke's account seems unable to explain the very normal occurrences of rewarding people for actions they might not be able to actively recall. We can therefore see that Locke's approach struggles to meet several features of identity, including the feature of survival which we witnessed in our discussion earlier in the context of the Alzheimer's patient that despite significant memory loss, we still consider (logically and intuitively) the person and individual as having survived. We can therefore conclude that Locke's memory criterion fails to achieve what a theory of personal identity ought to accomplish.

Derek Parfit and Psychological Continuity

It cannot be contested that Locke's theory of personal identity plays a great role in having influenced the work of Parfit. Building on the psychological criterion and concept of memories which Locke proposes, Parfit goes more in depth. Where Parfit deviates from Locke is by not depending as heavily on memories as Locke does, but instead emphasizing the role of what he calls *Relation R.* Parfit offers a view that better explains continued existence and demonstrates the connection between our present self and our "past selves" by suggesting that what is necessary for a person to continue as the same person over time is possessing *Relation R,* or both *psychological continuity and connectedness* (Parfit, 1984). Here, the concept of psychological connectedness refers to "the holding of direct psychological connections," (Parfit, 1984: 206) and rather than direct memories, includes *links* of memory and other mental events such as "continuing beliefs, desires, intentions, emotional attitudes, character dispositions, and so on" (Beck, 2015: 316). The concept of psychological continuity is "the holding of overlapping chains of *strong* (psychological)

connectedness" (Parfit, 1984: 206). Parfit (1984: 206) expands his description of these concepts as follows:

> *Of these two general relations, connectedness is more important both in theory and in practice. Connectedness can hold to any degree. Between X today and Y yesterday there might be several thousand direct psychological connections, or only a single connection. If there was only a single connection, X and Y would not be, on the revised Lockean View, the same person. For X and Y to be the same person, there must be over every day enough direct psychological connections. Since connectedness is a matter of degree, we cannot plausibly define precisely what counts as enough. But we can claim that there is enough connectedness if the number of direct connections, over any day, is at least half the number that hold, over every day, in the lives of nearly every actual person. When there are enough direct connections, there is what I call strong connectedness.*

In his *Reasons and Persons*, Parfit explains that *Relation R* is whatever matters in survival and thinks that what matters is psychological continuity and connectedness. Put differently, a person possesses psychological continuity and connectedness with an earlier one when *Relation R* is realized and is the same person over time when *Relation R* is present. *Relation R* is established when there is a chain of experiences, characteristics, intentions, values, beliefs, and so on, in a way that creates sufficient overlapping connections between the person we see today and that person from a time before (1984: 206). These overlapping connections refer to concepts such as memory links, sustained beliefs, and values, and so on, and when there are enough of these overlapping connections, *Relation R* is established (1984: 206; Beck, 2016: 117). A person is therefore one and the same person over time if they have *Relation R* where they possess sufficient direct (daily) psychological connections between their present and past self (1984: 206). "*Strong"* psychological connectedness is realized when there is sufficient overlapping of memories, maintained beliefs, and so on (Beck, 2015). Parfit describes sufficient connectedness as follows: "if the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person. When there are enough direct connections, there is what I call *strong*

connectedness" (1984: 206). Without sufficient overlapping mental and psychological chains and when the connections are really weak or non-existent, then we no longer have the same person. For Parfit, then, personal identity is a matter of degree, which leads us into the discussion of Successive Selves.

Successive Selves

When our grandmother tells us a silly story about the mess we made in the kitchen all those years back when we were 6 years old, we recognise this story as a retelling of an action and experience of our past self. Similarly, as I sit here writing, my mind drifts to envisioning what my future self, future Candice, will achieve within the next 5 years. With defending the position of personal identity being a matter of degree, Parfit believes that it makes sense to talk of earlier or later selves of a person. Parfit (1984: 302) introduces the concept of *Successive Selves,* an extension of his psychological account which he uses in an attempt to describe the interrelations between a person and those persons who are like them, where each of these people would be a 'self' that indicates a respective psychological distance compared to the self and person today, namely the *present self* (1984: 302). According to Parfit, earlier and later selves are clearly not *exactly* like each other. Instead, the differences will be so great that it is *as if* they are someone else but are technically still the same person in being uniquely psychologically continuous. Parfit explains that because psychological continuity functions in both directions of time, a *self* includes talk of both an *ancestral self* and a *descendant self* (1984: 302). Parfit suggests that it is naturally intuitive to speak of what we consider our past (ancestral) or future (descendant) selves.

Parfit (1984) goes on to discuss how different selves of a person could be a way of describing the different degrees of psychological connectedness and that this functions in the way of giving a new meaning to the concepts of *my past self* and *my future self*. When speaking of these selves, we may feel a sense of disconnect. While we recognise that the 6-year-old who created a mess in the kitchen *was me* when I was younger and I acknowledge the connection between myself today and my past

distant self, I no longer feel like the same *person* as I once was. Parfit uses the way we speak here to explain that the terms of past or future self do not refer to myself or to different people, but rather those persons who share a relation with me (my present self) by means of a degree of psychological connectedness (1984: 302). In other words, when speaking of a past self (such as the young girl of 6), I am speaking about something who shares a degree of psychological connectedness to my present self. When speaking about one of my past selves from, suppose, four years ago when living abroad, I feel a stronger connection to this person as "one of my closer past selves" (Parfit, 1984: 302). Speaking of a closer past self suggests that the degree of (psychological) connectedness between my present self and this specific past self is stronger than the connection between my present self and my past self of when I was 6 years old, which would be a more distant self. Thus, whether we speak of a close past self or a more distant past self, we are indicating a degree of psychological connectedness between the person we are today and the person we were at a different stage prior to today. The concept of *Successive Selves* is used as these different selves would typically work in succession from our most ancestral self with whom we feel the least connected to, to a less distant self we share a slightly stronger degree of connection to, up to the closest self we feel the strongest degree of connection to and ultimately the self we recognise as our present time. It should be noted, as Parfit mentions, that this explanation of past-directed series of phrases of oneself can be used in the same way for a future-directed series.

Parfit offers the case of *The Nineteenth Century Russian* to further demonstrate his concept of Successive Selves (1984: 327):

> *The Nineteenth Century Russian.* In several years, a young Russian will inherit vast estates. Because he has socialist ideals, he intends, now, to give the land to the peasants. But he knows that in time his ideals may fade. To guard against this possibility, he does two things. He first signs a legal document, which will automatically give away the land, and which can be revoked only with his wife's consent. He then says to his wife, 'Promise me that, if I ever change my mind, and ask you to revoke this document, you will not consent.' He adds, 'I regard my ideals as

> essential to me. If I lose these ideals, I want you to think that I cease to exist. I want you to regard your husband then, not as me, the man who asks you for this promise, but only as his corrupted later self. Promise me that you would not do what he asks.'

Parfit uses the above scenario to demonstrate the credibility of the concept of Successive Selves where we intuitively consider a version of ourselves that shares more connections to our present self as being more closely related to the person we consider ourselves to be. This allows for the natural changes we see in a person's beliefs, values, and so on, but once these changes become so great, it is not unreasonable for us to consider that while this is not a completely new or different person, this changed version of ourselves is instead a distant self with whom we share fewer or weaker connectedness, *as if* they are someone else.

Teletransportation

Let us turn back to one of Parfit's most frequently cited thought experiments – teletransportation, to examine his psychological criterion (1984: 199), although it is worth noting that the thought experiment as quoted below are not Parfit's own words:

> *Suppose there exists a device, a teletransporter. One on Earth, and one on Mars. A man on Earth steps into the device which then scans each one of his cells in the exact state in which they are in that very moment. The device then teletransports all the perfectly scanned information about the man to the second teletransportation device on Mars, all while the original cells of the man on Earth are being destroyed. The outcome here is that an exactly identical version of the man on Earth is created out of new matter in the device on Mars.*

The person who steps out of the teletransporter on Mars is exactly like the original man on Earth in the sense that he thinks like the original man and possesses all the same memories, beliefs, and even physical scars as the man who stepped into the teletransporter on Earth. In fact, the psychological states of the replica on Mars are the direct casual result of those of the Earthling, not simply similar ones. It would not be unreasonable to consider what has just happened as a form of

long-distance travel (Parfit, 1984: 200) instead of a person on Earth being destroyed and another coming into existence on Mars that just happens to be like to original person on Mars. Parfit considers the being on Mars as the survived person as there is enough of the original person that has survived to maintain sufficient psychological continuity and connectedness between the Earthling and the being on Mars.

An interesting challenge arises in the scenario where the being on Earth was in fact not destroyed. Let us call this a case of accidental replication. Known as the Branch-Line Case (1984: 201), the being on Earth has his body damaged in such a way that he will meet his demise within a few short days, all the while the experiment is a partial success and a being is created and steps out of the device on Mars with a body that is exactly like mine at this very moment (although constructed of different matter) and psychologically continuous and connected to the being on Earth who entered the machine (Parfit, 1984: 201). Parfit is confronted with the challenge of being forced to confirm that both the fatally damaged Earthling and the Martian are one and the same person, as there exists *Relation R* between them. However, this proves difficult to accept. For many, if not most, intuition would be leading us to believe that the Earthling is the original person, and that the Martian is simply a *replica* of the (original) Earthling. Instead of a case of long-distance travel, it would appear that we are now being confronted with a case of accidental *replication* where the replica is just that – a replica – and *not* one and the same as the original Earthling.

Recall that in the simplified version of this scenario (where the Earthling is destroyed and a new being steps out of the device on Mars), we accept and agree that the Earthling has survived as the being on Mars. Suppose while the being on Mars stepped out of the device, as did the original person out of the device on Earth, but not to our knowledge. We would be of the understanding that the person we knew on Earth has travelled to Mars. However, upon discovering that the original Earthling has survived, whether for the next week or the next decade, doubt would be cast over the identity of the Martian, as we would revert back to believing that the being on *Earth* is the original

person, and *not* the being on Mars. The concern raised here is that it appears that the existence of a separate, external being could after all influence the identity of another. Put differently, it seems absurd to suggest that the identity of the person who stepped out of the device on Mars depends on whether or not we *know* that the original Earthling survived. If the Earthling did not survive or we were not aware of their survival, we consider the being on Mars to be the surviving Earthling who simply travelled across an extraordinary distance. On the other hand, if the Earthling does survive, we strongly consider the person on Earth as the original surviving person, and the being on Mars simply a replica or clone of the Earthling, but that they are *not* one and the same as the Earthling. Although according to Parfit's psychological criterion both the fatally damaged Earthling and Martian are psychologically continuous and connected to the original Earthling, neither the fatally damaged Earthling nor the Martian are psychologically closer than the other to the original Earthling. However, although Parfit's *Relation R* is ultimately psychological continuity and connectedness, we cannot ignore the fact that the dying Earthling has the original body (of the Earthling) and this is what intuitively makes them (and not the Martian) identical to the original Earthling, a serious charge Parfit is required to defend his psychological criterion against.

To further support this concern, let us consider the following scenario. In the simplified version (where the being on Earth has its body immediately destroyed), the surviving person on Mars commits a crime. We would find it reasonable and unquestionable to hold the Martian accountable for their actions and punish them accordingly. Likewise, if the device on Mars malfunctioned and the Earthling stepped out their respective device unharmed and went on to commit a crime, we would expect the person on Earth to be held responsible and prosecuted. In the case where both the person on Earth and on Mars survive, suppose the Martian commits a crime. We would naturally argue that the Martian should be held fully accountable, not the person on Earth, demonstrating that we do in fact consider these two beings as different from one another, and that it is *not* the case that the original Earthling simply travelled to Mars. Consider this scenario taken one step further. Suppose the Earthling is not aware of the existence of the being that stepped out of the

device on Mars, and they (the Martian) go on to commit a crime (on Mars). It seems rather odd to suggest that should a replica of ourselves be made, especially without our knowing, and if they commit a crime that we would and even *should* be held responsible for their misdeeds. The Earthling would attest that *they* did not commit the crime. Intuition tells us that this is rightfully so – the being on Earth did not *physically* commit the crime. It was the being on Mars who committed the crime, while under no control by the Earthling. There are no psychological connections (and therefore no causal relations) between the two (the Earthling and the Martian). They are simply psychologically similar. Regardless of the psychological similarities between the Earthling and Martian, it is not the case that the Earthling makes decisions that are enacted by the Martian, and vice versa, as there is no causal relation. After each being steps out of their respective devices, regardless of the level of *Relation R,* intuition tells us that the one cannot be held accountable for the other's actions as if they were exactly one and the same, and that if it were the case that one was to commit a crime, then it should be *that* physical version of the being be held responsible for their misdeeds, not simply either of them. All the scenarios that have been discussed in this section contribute to concern about the plausibility of Parfit's psychological view and his *Relation R* criterion where not only do some cases appear questionable, but nonsensical and counterintuitive, too.

One may immediately, after reading through the above synopsis of Parfit's PCT (psychological continuity theory), accuse Parfit of failing to provide a credible explanation of personal identity. This would be a justified concern as under Parfit's criterion of *Relation R,* it seems that it is possible for a person at T_1 to not be the same as person at T_2, despite appearing to still be the same person. A fundamental point where Parfit differs not only from Locke but also from other theorists is what he considers to be important to personal identity. Parfit asserts that the concept and logic of personal identity is in fact *not* what matters when it comes to continued existence (1984: 206). When *Relation R* is present, namely sufficient psychological continuity and connectedness, the person, according to Parfit, *survives,* and it is this psychological continuity which Parfit believes matters *more* (1984). Let

us turn to Parfit's dispute with non-reductionism and his *My Division* argument (reduplication) to understand and make sense of the concepts involved here.

Reductionism and Non-Reductionism

It is worth pausing here to provide a brief discussion around the dispute between reductionism and non-reductionism and how they differ with regards to what matters in survival. Parfit (1984: 210) describes *Reductionism* as a person's identity being a matter of "certain more particular facts" (such as a physical or psychological continuity). In other words, personal identity can be 'reduced' to a simple explanation – such as physical or psychological continuity. What matters to survival is one of these facts – physical or psychological continuity, for instance. On the other hand, *Non-Reductionism*, as described by Parfit (1984: 210), does not accept that identity comes down to physical and/or psychological continuity. Instead, it involves "a further fact" (Parfit, 1984: 210). Parfit (1984, 210) describes this *Further Fact View* of non-reductionism as insisting that identity is a (further) fact of its own kind. This view stipulates that we exist along with these facts, but that "personal identity *is* a further fact, which does not just consist in physical and/or psychological continuity" (Parfit, 1984: 210), and that personal identity cannot be 'reduced' to a basic fact such as physical or psychological continuity. Non-Reductionists would argue that basic facts such as a physical or psychological criterion, are not what ultimately matter in survival, and that it is the *further fact* that matters in survival. As we continue to see below, Parfit attempts to use his Reductionist approach and *Relation R*, rather than a Non-Reductionist view, to explain personal identity, and that is this *Relation R* that matters in survival.

Before looking at Parfit's case of *My Division,* consider the following case (1984: 253):

> *Suppose first that I am one of a pair of identical twins, and that both my body and my twin's brain have been fatally injured. Because of advances in neuro-surgery, it is not inevitable that these injuries will cause us both to die. We have between us one healthy brain and one healthy body. Surgeons can put these together. This could be done even with existing techniques. Just as*

> *my brain could be extracted, and kept alive by a connection with an artificial heart-lung machine, it could be kept alive by a connection with the heart and lungs in my twin's body. The drawback, today, is that the nerves from my brain could not be connected with the nerves in my twin's body. My brain could survive if transplanted into his body, but the resulting person would be paralysed. Even if he is paralysed, the resulting person could be enabled to communicate with others.*

Ignore for just a moment whether the physical surviving body is paralysed or not and assume that the procedure is a success and that my twin's body, containing my brain and all my mental contents, together have survived the operation. Parfit (1984: 253) believes the question of who the identity is of the surviving person is not difficult and explains that there is no reason for anyone (whatever their theory of personal identity) to deny that I (the 'I' in the scenario) survive in the single case as the resulting person will be psychologically continuous with me (and not my twin), regardless of any physical elements being involved. As Parfit puts more clearly, "if all of my brain continues both to exist and to be the brain of one living person, who is psychologically continuous with me, I continue to exist. This is true whatever happens to the rest of my body" (1984: 253).

We know through science that it is in fact possible for a person to survive with only one functioning hemisphere of their brain (to clarify, both the *person* and the human being survive, particularly in cases where a person has suffered a stroke or some other form of neurodegeneration that has affected only one half of their brain. Parfit (1984: 254) uses this knowledge to posit another potential outcome. First (and as demonstrated above), "I would survive if my brain was successfully transplanted into my twin's body. And I could survive with only half my brain, the other half having been destroyed. Given these two facts, it seems clear that I would survive if half my brain was successfully transplanted into my twin's body, and the other half was destroyed" (Parfit, 1984: 254). The question now is what would happen in a case where the other half on brother A's brain was *not* destroyed? This leads us to Parfit's *My Division* case (1984: 254), which is as follows:

> *My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people*

> *believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine.*

Knowing that it is possible to survive with only one functioning hemisphere, but not yet having the technology to divide a person's brain in half and transplant each half into a different body with the survivor being physically unaffected is, according to Parfit (1984), a mere technicality and that because we know a person *can* survive with only one functioning brain hemisphere, we can assume that both brothers have survived when each receiving one half of brother A's undamaged brain. In this scenario, not only are both the two brothers psychologically continuous with me, but both are *equally* psychologically continuous with me. Both being (equally) psychologically continuous to me means they are, at least for a moment, psychologically continuous to one another. It is here that Parfit (1984: 255) points out what he considers as *not deeply* impossible in this scenario – a person's consciousness being divided into two separate streams (citing the evidence from commissurotomy patients). However, it would be nonsensical to suggest the brother A has survived in both the bodies of brother B and body C as logically, the *same* thing cannot survive in two places simultaneously. But, with both brothers being psychologically continuous with brother A, we cannot argue that one (either brother B or brother C) has survived as brother A and that the other brother has not. In other words, we cannot claim that brother A survives in brother B's body and *not* brother C's body, and vice versa, because they are both equally psychologically and physically continuous to brother A. Additionally, recall that we have no difficulty accepting that brother A survives in brother B's body if the other half of A's brain is destroyed. But it seems somewhat troubling to suggest that the identity of brother B's body depends on whether brother C's body with the other half of A's brain survives successfully.

While Parfit rejects the importance of a physical criterion, it is interesting to consider how the physical body of the recipient (and its appearance) plays a role in who we think survives. For instance, we have an easier time believing that brother A survives in brother B (or C's) body as

opposed to if brother A's brain (or one hemisphere) was transplanted into a completely different-looking body. Parfit (1984: 254) offers an example to demonstrate this:

> *It would be odd for a six-year old girl to display the character of Winston Churchill, odd indeed to the point of outrageousness, but it is not utterly inconceivable. At first, no doubt, the girl's display of dogged endurance, a world-historical comprehensiveness of outlook, and so forth, would strike one as distasteful and pretentious in so young a child. But if she kept it up the impression would wear off.*

Again, although Parfit discounts the physical criterion, the above example illuminates the possibility that our assumptions with the original *My Division* case are distorted as a result of the very close physical similarity between the three triplet brothers. Perhaps it is easier to view brother A surviving as brother B or C *because* the resulting person will not only be psychologically continuous with brother A, but also look identical to him. This point, however, will not be discussed in more detail and was simply included as a point of interest.

Parfit's (1984: 255) remedy to the problem of the *My Division* case is his conclusion that "personal identify is not what matters" in survival. Parfit uses the range of possible outcomes of *My Division* to demonstrate that personal identity is not what matters in survival. Although some were mentioned above, let us consider all the options here more closely. Parfit (1984: 256) offers the four following potential outcomes, namely, (1) I do not survive; (2) I survive as one of the two people; (3) I survive as the other; (4) I survive as both where the relation between each of the survivors and the original in *My Division* is exactly the same as the relation between the survivor and the original in the single case. Looking at the first possibility, the objection would be that we know that we *can* in fact survive with one hemisphere, so this cannot be the outcome. Then, the objection against the second and third possibilities would be what was mentioned just moments ago – as both brothers would be equally psychologically continuous to me, there is no way for us to argue that brother A survives as one of the brothers instead of the other. Finally, Parfit (1984) argues that although the fourth outcome is possible, it is possible in a way that is not congruent with our notion of identity where

the necessity of identity is described as when only one thing can be identical with any other thing. This is because what we have in *My Division* is a case where there is everything that matters in survival (as revealed by each half being the same as the single case), but in which there is not identity because there are *two* of them.

Schechtman raises a significant concern worth covering briefly. In the single case where one of the brothers wakes up with brother A's consciousness, we are more prepared to accept that brother A has survived in brother B's body, as the consciousness in brother B's body is identical to brother A's consciousness prior to the procedure. However, upon learning that brother C has also awoken with brother A's consciousness (due to possessing the other half of brother A's brain), we now question who has woken up in brother B's body. Schechtman points out that whether brother A is identical to body B (the original brother B) seems to depend on whether body C (the original brother C) survives or not. Should brother C die, brother B would suddenly be identical to brother A. That is, on Parfit's account, identity is no longer necessarily an *intrinsic* matter. Parfit is prepared to accept this consequence on the ground that it affects all reductionist views and only a non-reductionist view in which we are separately existing entities like souls is exempt. However, since such a view is false, we must live with the possibility of identity being extrinsic. Schechtman (1996), however, argues that this is completely counterintuitive and works against Parfit.

For Parfit's psychological criterion to be suitable, it ought to satisfy the four features of identity. First, let us consider the feature of survival. As discussed in great detail, we would intuitively argue that a person continues to survive regardless of the psychological changes they may experience. Even in cases where a person suffers from total amnesia and possesses no psychological continuity or connectedness to their previous selves, we would not say the person has not survived. Of course, we may speak in a way saying that the person we once knew is no longer the same or that we feel like the person we once knew no longer exists, but we do not literally mean that the *person* no longer exists. Consider my amnesiac grandmother. Although she may have no recollection of any

moment in her life, I still recognize my grandmother as having survived, but instead with total amnesia. For now, Parfit's psychological criterion is unable to satisfy the feature of survival.

The arguments for the features of moral responsibility and compensation are similar to what was discussed with Locke. Even in cases where I have no recollection of insulting my friend, this does not take away the anger that is being directed towards me. I naturally feel the need to apologise and make amends, and although I have no memory of hurting my friend, I am still seen as (literally) the same person who caused the hurt as there is continuity between my current self and this past insulting self. Simply having forgotten about the insult would not free me of being responsible because according to Parfit, I am still the same person as I am continuous (even if not connected to – by not possessing the specific memory of the insult) that self from the past who insulted my friend. Likewise with the feature of compensation. Regardless of whether I have any psychological connection to the person I was when I saved a neighbour's kitten from being run over, it would not be unreasonable for the neighbour to still want to thank, and even reward me, the person I am today. Imagine a retired actor has no memory of filming a particular movie when they were younger. It would be absurd to withhold the royalties (money made from the film, for example) from the actor simply because of this memory loss. We would continue to compensate the actor for their hard work many years ago, albeit no longer being psychologically connected to that version of themselves, as there remains *continuity* between the actor today and their self from before.

The fourth feature, self-interested concern, is the final feature with which Parfit's criterion struggles. As we will later see in our discussion of Williams, being told that I will not remember the pain I will experience tomorrow does not make me fear this pain any less. If I am told that I will be in a car accident next week, but I will have no memory afterwards, I will not feel any less sense of impending doom and I would be plagued with thoughts around my survival from the accident. I would be thinking about whether *I* would survive the accident, what would happen to *me*, what kind of life *I* would live after the accident, regardless of losing all my memories afterwards. Here we witness that regardless of whether I possess my memories, I still feel a sense of egoistic concern, a

special type of concern for *my* future and *my* survival, ultimately exposing the trouble Parfit's psychological criterion faces in terms of satisfying this feature of identity.

Jeff McMahan and the Embodied Mind View

McMahan does not exclusively subscribe to a physical or psychological approach to personal identity. Instead, in *The Ethics of Killing: Problems at the Margins of Life* (2002), McMahan puts forward his Embodied Mind Approach of identity where he places significant emphasis on the need and requirement for physical continuity for persisting personal identity; however, also argues that physical continuity alone is not sufficient for continued personal identity as he does not, as mentioned, consider personal identity to be strictly due to a physical criterion. What is meant by not fully accepting just a physical criterion is that unlike traditional animalists, for McMahan (2002: 50), a being simply possessing their brain does not constitute personal persistence. Rather, McMahan proposes that identity is the result of "the continuity of the capacity for consciousness" (McMahan, 2002; Oyowe, 2010: 275). McMahan explains that for the continuity of the capacity for consciousness, there needs to exist continuity of the same physical basis of consciousness (Oyowe, 2010). Continuity of the capacity of consciousness is seated, according to McMahan, in the *same brain* (Oyowe, 2010). It should be mentioned here that Oyowe's interpretation (2010) of McMahan is being used here to support the concerns I share regarding the emphasis McMahan appears to place on a physical criterion. To emphasize, simply possessing a brain is insufficient as the brain could, for example, be in a vegetative state and not be functional. Rather, continuity of the same physical brain is required, and identity is maintained if the same brain maintains *functional continuity* that supports the *capacity* for consciousness (Oyowe, 2010: 275). It is in stressing the *capacity* for consciousness that signals where McMahan deviates from traditional animalists. Without this capacity for consciousness, McMahan would deny that identity continues. Importantly, the capacity for consciousness on its own is also not sufficient for persisting personal identity,

according to McMahan (2002). It is necessary for persistence of the same physical brain that supports the capacity of consciousness that leads to continued identity.

Although one may interpret McMahan as advocating a psychological criterion, this would be erroneous. McMahan does not concern himself with the *mental content* of a person's psyche and even goes as far as to reject the importance of psychological connectedness and continuity (Oyowe, 2010). This unveils a crucial point where McMahan differs from Parfit. McMahan (2002) acknowledges that many people share the same intuitive response as Parfit in the case of *Teletransportation,* where we consider having survived as the replica on Mars. McMahan (2002), however, does not share this intuition, and finds trouble in the fact that the existence of the person on Earth interferes with the identity of the replica on Mars. In the case where both the original earthling and the replica exist and both possess exactly identical psychological continuity and connectedness with the original earthling, as we saw earlier, Parfit concedes that both the original being and the replica are one and the same. McMahan states that "the original person has as much reason to care egoistically about the replica as an ordinary person has to care about herself on the following day" (2002: 57). McMahan argues that Parfit's intuitive response of who survives is misplaced. This is where we see McMahan's Embodied Mind View differ from Parfit. This divergence can be made clearer with McMahan's (2002: 57) scenario of *"The Suicide Mission"* which is as follows:

> *The Suicide Mission. In a time of war, one has been chosen to carry out a military mission that will involve certain death. Although the operation of the Replicator is very expensive and has therefore been strictly rationed, one's superiors have granted one the privilege of having a replica of oneself made prior to the mission. They will also allow one to choose, prior to the process of replication, whether one will go on the mission oneself or whether the replica will be sent. (Because one is a dutiful soldier, one's replica will be dutiful as well. One knows that if ordered, he will go on the mission.)*

McMahan believes that we would, without hesitation, choose to send our replica on the mission (2002: 57). This assumption is seated in the consideration of egoistic concern. It is that because I care about what will happen to me in the future, and I do not want to volunteer myself for certain death, I will send my replica and avoid this inevitability for myself. McMahan (2002) uses *The Suicide Mission* to illuminate the troubles with the Psychological Account, particularly Parfit's psychological approach. If what is necessary when discussing personal identity is *Relation R* - enough psychological continuity and connectedness – then in the above scenario, the replica would be psychologically continuous and connected to the original soldier. This would mean that it would make no difference whether the original soldier or the replica is sent on the suicide mission to meet their inevitable death (2002: 57). To detail this problem more explicitly, if the soldier and replica are one and the same and it does not matter who is sent on the mission, then an acceptable option would be to have the original soldier be sent on the mission and meet his certain death, while the replica stays behind and continues living the original soldier's life. However, this does not seem to align very well with our natural intuition. It seems counterintuitive to suggest that we would, ourselves, go on the mission and meet our guaranteed death, but that we will survive as our replica. Similarly, if we were to send our replica on the mission, it is not unlikely for many of us to consider that it is our *replica* and not *ourselves* who perishes. Furthermore, if the replica is psychologically connected to and continuous to the original soldier, then the original solider would be concerned (egoistically) about the replica in precisely the same way that they are concerned about their own future. However, there appears to be something troubling about the soldier being *as concerned* about the replica as they are about themselves. It seems unlikely that the soldier will consider the replica as one and the same as himself, let alone volunteer themselves to go on the suicide mission *in place of* the replica. For most, it is not surprising to consider the replica going on the suicide mission as a ticket for the soldier to avoid a certain death. In a scenario quite like Parfit's case of Teletransportation, specifically the version where the original body is in fact not destroyed, we see a justified worry from McMahan about the plausibility of Parfit's psychological approach.

Interestingly, however, McMahan discusses another case that seems to contradict our intuitions regarding *The Suicide Mission*, and in fact offers support for Parfit. I bring this to attention as not doing so would be omitting important considerations and falling into the trap that many others do – using only the arguments/scenarios that serve a particular narrative. "*The Nuclear Attack"* (McMahan, 2002: 58) is explained as:

> *One is an employee at the Pentagon, which has a Replicator capable of transmitting one's cellular blueprint to a replicating booth in Alaska. One receives confirmation that a nuclear missile, targeted on the Pentagon, has penetrated the country's defences and will obliterate the entire area within a minute. That is just enough time to have oneself scanned and for the data to be transmitted to Alaska.*

McMahan admits that we would naturally opt for a replica to be created in the Replicator in Alaska compared to our original selves being destroyed and no replica being created at all, and that this type of reduplication will, in a sense, off all – not part – of what matters for survival (2002: 58). McMahan (2002) even goes as far as to say that in this case of reduplication, it would be a *mistake* to consider this a risk to our survival, because we will not know until after the procedure what will happen and if whether what matters for our survival has been preserved. However, this position differs significantly from what we witnessed just moments ago in *The Suicide Mission.* McMahan (2002) concedes that because this case is not consistent with the physical criterion, nor is it in line with our intuition about what will happen, we cannot view this case as a clear instance of survival. Oyowe (2010) brings our attention to a damaging concern that McMahan faces here, namely that there is no formal argument provided by McMahan to support his claim that *The Nuclear Attack* is not a case of survival. McMahan explicitly discusses and denies that we would have any – or at least the same level of – egoistic concern for our replica in *The Suicide Mission*, indicating that McMahan thinks that the replica does not possess what matters for survival (Oyowe, 2010). However, McMahan fails to offer the same attention to explaining why we cannot consider *The Nuclear Attack* as a case of survival.

One final scenario I would like to offer here is McMahan's discussion about *The Cure* (2002: 77):

> *Imagine that you are twenty years old and are diagnosed with a disease that, if untreated, invariably causes death (though not pain or disability) within five years. There is a treatment that reliably cures the disease but also, as a side effect, causes total retrograde amnesia and radical personality change. Long-term studies of others who have had the treatment show that they almost always go on to have long and happy lives, though these lives are informed by desires and values that differ profoundly from those that the person had prior to treatment. You can therefore reasonably expect that, if you take the treatment, you will live for roughly sixty more years, though the life you will have will be utterly discontinuous with your life as it has been. You will remember nothing of your past and your character and values will be radically altered. Suppose, however, that this can be reliably predicted: that the future you would have between the ages of twenty and eighty if you were to take the treatment would, by itself, be better, as a whole, than your entire life will be if you do not take the treatment.*

Should you take the cure? Would it be rational and in your best egoistic concern to take the treatment? McMahan is under the impression that most people will be sceptical about the treatment, while many others would feel strongly against taking the cure (2002: 77). McMahan uses this hesitation to offer support for a psychological criterion, much like that we see Parfit offer. As McMahan (2002: 78) puts it, "The future you would have with the treatment would contain vastly more good than you will have if you refuse the treatment, but the future offered by the treatment is too much like someone else's future. In that future, you would be a complete stranger to yourself as you are now. The psychological distance between you now and yourself as you would be after the treatment is too great for you to think of the goods in that future as fully *yours.*" This psychological distance is seemingly what causes our hesitation of taking the cure as we do not expect that it will be *us* who survives, meaning that what in fact matters for survival is not preserved within a psychological criterion. If we have no psychological continuity to this future stranger and anticipate no psychological connection, it follows that we would be hesitant in taking the cure as this would be in the best egoistic concern of someone *else's* future, not my own.

Looking at how McMahan's Embodied Mind Account fares against *The Cure*, we find some trouble for McMahan. Recall for a moment, McMahan's approach. What is responsible for personal identity is the maintenance of one's functioning physical brain that supports their *capacity* for consciousness. In *The Cure,* I (the 20-year-old diseased person) experience no physical changes, and with no changes to my physical body, particularly my brain, it is expected that I still possess the functioning (and physical) capacity to support my consciousness. McMahan (2002: 78) postulates that if his Embodied Mind Account is accurate, then it will simply be the case that becoming a total stranger from who I see myself as today is just what is in the cards for my future. However, as McMahan fairly points out, myself today would not be sufficiently connected to this future person, so it would not make rational sense for me today to care about this future person in the same way that I care about *my* future.

My concern here, however, is that McMahan has not adequately addressed the challenge his Embodied Mind View encounters. It appears McMahan is diverging slightly and seems to use a psychological criterion (much like what Parfit offers) to explain our intuitive response against taking the cure. However, McMahan has not satisfactorily dealt with the concern discussed above regarding his own view. For now, McMahan does not provide a clear defence why we intuitively feel against taking the cure with the anticipation that *we* will not be the ones who survive. Again, we are physically unharmed, and the entirety of our brains will remain intact and untouched, therefore supporting the capacity for our consciousness. So, it is curious as to why irrespective of these criteria being met, we are *still* somewhat hesitant (and even against) taking the cure. McMahan, at least for the time being, fails to offer some sort of resolution to this concern.

Suppose McMahan's retort would be that his Embodied Mind Approach requires one's physical brain the support (the capacity of) *their* consciousness, and not simply *anyone's* consciousness. This way, it follows that *someone* will exist after taking the cure. However, carefully consider what is being said here. If this defence were to hold, there would already exist an assumption of whose consciousness the brain is supporting. If this is the case, then we are forced to admit that it is not the

physical brain and capacity for consciousness that carries personal identity, but instead that there is something else responsible for personal identity if we are speaking in a way of *our* consciousness, our physical brain supporting *our* consciousness, and what will happen to *our* consciousness after taking the cure. If it were the case that McMahan offered a response along these lines, his view would fare no better than it currently does.

We move on to assessing how McMahan's criterion fares against the four features of identity. Recall McMahan's criterion – requiring sufficient continuity of the physical brain in order to facilitate mental functioning and the capacity for consciousness. Of course, when a person possesses their physical brain but suffers severe memory loss (although maintains basic mental and cognitive functioning), we still consider them the same, surviving person, and this fits neatly with McMahan's criterion where the requirement of one's physical brain that supports functional continuity and the capacity for consciousness appears sufficient for one's survival. However, the trouble comes in when studying various thought experiments. Consider again *The Nuclear Attack*. A missile is on its way to annihilate the building I am working in, but I have the option of creating a comprehensive blueprint of myself and my molecular makeup and having a replica of myself constructed in a device in Alaska. As covered earlier, we would interpret this as a form of survival. However, there is no continuity of my physical brain. In fact, my original brain was destroyed while a new physical version of myself was being constructed elsewhere, meaning that there is no physical continuity between my original brain, and this replicated brain. But it is not unreasonable to be inclined in believing that I survived in Alaska. Cases like this reveal some trouble McMahan faces.

Before moving on to the next feature, it is interesting to take note of another difficulty with McMahan's criterion and the feature of survival. Suppose I suffer from a degenerative brain disease that would lead to my demise in 2 years, but my mental functioning and capacity for consciousness remains unaffected. I am made aware of a brain transplant procedure where I will receive a donor's brain (of a person who died of a sudden heart attack). Prior to the procedure, my brain is scanned, and all my mental content is perfectly copied into the new brain at the end of the surgery. I wake up

and possess all the same memories I had before the operation and feel as though nothing has changed, apart from being in immense physical pain. The problem McMahan faces is clarifying whether continuity of *my* physical brain is required, or whether *any* physical brain is sufficient, as long as continued functioning and consciousness is maintained. My argument here is that McMahan has not provided enough clarity and specificity in his criterion. So, the question we are left with is whether I believe that I survive because *my* consciousness is supported by any physical brain, or whether I do not survive because there is no continuity of *my* physical brain, an opinion which works against our intuition. With these questions unable to be answered at this time, McMahan finds himself in the scenario where this problem does not only affect the feature of survival, but other features, too.

We see the same problem discussed above spill over into the remaining features – whether we require continuity of our own brain, or if continuity of consciousness in any physical brain will suffice. Let us, for now, assume we need continuity of our own brain as this avoids any interruption in this physical continuity, while also ignoring that there exists a glaring need of a prerequisite condition for it to be considered *my* brain. Possessing continuity of our brain that maintains functioning and consciousness (irrespective of which memories, if any, are intact) aligns well with our assumption that we have survived, and it is our survival demonstrated here that follows neatly in me being held morally responsible or deserving of reward. Seeing as I maintain mental functioning and my consciousness, there is no reason I would not be held responsible for previous actions or compensated for my good deeds. This works in McMahan's favour. However, it becomes less clear in cases where psychologies are swapped between bodies. If my physical brain and body are left untouched but a swap between my and my friend's psychologies takes place, it might appear as if I have woken up in my friend's body and vice versa. If this is in fact the case, then the continuity, and even location and possession of a particular brain plays no role in personal identity and it would therefore follow that if I, in my friend's body, commit a crime, we would find it difficult to want to hold my body – with my friend's psychology – accountable. Similarly with compensation. If I found

the cure for cancer, albeit in my friend's body, it is only just that I am rewarded. It seems absurd to reward my body, a mere vessel, especially one that appears to behave like someone else entirely and who (the body and psychology) had no role in my finding of a cure. It therefore seems that if McMahan is requiring the continuity of one's own physical brain, that this is insufficient for personal identity and for satisfying these features of identity, never mind the additional challenge McMahan faces when needing to explain how we can see it as *my* brain.

Turning finally to self-interested concern. Let us imagine the same procedure above has taken place and my friend and I have our psychologies swapped between our two bodies. After waking up and realizing I feel pain from the operation but in my friend's body, I would be concerned about the future *I* will live in my *friend's* body. I will be concerned about how *I* will recover and adjust to life in my friend's body. If one of us were to be harmed, I would fear harm towards me while in my friend's body more than I will fear my actual body being harmed, which is now seemingly disconnected from me. It seems absurd to suggest that if I believe I have woken up in my friend's body, that I would be more concerned about what happens to my body instead of where I think I have survived. We again see some trouble McMahan faces as his criterion insufficiently satisfies the feature of self-interest.

Physical theories of Personal Identity

Bernard Williams and Bodily Continuity

Aside from psychological theories of personal identity, several philosophers have considered significantly different notions of personal identity, namely that of physical (or biological) continuity, where a person is one and the same person over time if there is continuity of their physical body. One of the central contributors to the physical continuity theory defence of personal identity is 20th century philosopher, Bernard Williams.

Williams rejects the psychological criterion of personal identify and instead offers an explanation that differs considerably from that of Locke and Parfit, although he (Williams) uses a modified version of Locke's soul-swap (the Prince and the Cobbler) thought experiment to defend his position (Williams, 1970; Beck, 2016). While Locke discusses a soul-swap, along with a transfer of psychology, Williams argues that a transfer of psychology does not have to be seen as a *person* transfer. Williams (1970) does not use a soul, as Locke does, but a 'brain-state transfer device' to move the psychology. Williams posits that it would not be unreasonable for someone to view the scenario as a person remaining in their own body, but simply receiving a new psychology (Beck, 2016). The modified version presented by Williams (1970: 161-163) postulates that *Person A* has all contents of their consciousness (memories, beliefs, and so on) scanned and then copied into *Person B*'s body who had moments before had their entire consciousness erased. Expecting that *Person B* will now begin behaving in a way that is recognizable with the thoughts and actions of *Person A,* Locke would, as demonstrated earlier, conclude that *Person A* has awoken (or survived) in *Person B's* body, or that *Person A* now occupies *Person B's* body. Williams (1970) vehemently disagrees with Locke and instead suggests that it is *Person B* who has survived, not *Person A,* regardless of the extensive changes to their respective psychologies. Williams (1970) rejects the need for continuity of the same consciousness and instead insists that because *Person B's* body continues to exist, it is in fact *Person B* who survives, albeit with a significantly different consciousness. For Williams, regardless of how a person thinks or behaves, even if considerably different from their usual thoughts and behaviours, if their body continues to exist, it is this continuity of their body that supports their survival and continued existence.

To further defend his view, Williams (1970) puts forward an alternative thought experiment, also a revised and modified version of the consciousness change originally seen in the beginning of this paper. Let us begin with Williams's second presentation as detailed in *"The Mad Scientist"* (1970). Suppose *Person A* is told that they will be tortured tomorrow; however, the experimenter explains to *Person A* that all their memories will be erased moments prior to the torture, and thereafter

replaced with new memories which belong to another person (Williams, 1970). However, regardless of any level of psychological interference they are told they will experience between now and the torture, it seems quite reasonable, and even expected, for *Person A* to report feelings of dread of the impending torture, proclaiming that their fear of the future torture has not been lessened by the news of the changes they will experience to their psychology. Williams uses this to support the claim that *Person A* still thinks *they* will be the one who will be tortured, irrespective of the considerable psychological changes they will undergo, including their psychology being replaced by *B's* psychology, revealing that *Person A* places significant emphasis, and even depends on bodily continuity for their own survival. *Person A* displaying compelling concern for the fate of *body A* as if it were their own is a fundamental insight on which Williams's physical criterion relies as it works to demonstrate his suggested view that a person can survive an entire psychological change granted that there is bodily continuity.

For the sake of completeness, let us cover Williams's first presentation in *The Mad Scientist* (1970). Imagine a scenario where *Person A* has all contents of their consciousness transferred into *Person B's* body, and *Person B* has their entire consciousness transferred into *Person A's* body (Williams, 1970: 167-178, as cited in Bennett 2014). Ahead of the transfer, the Mad Scientist informs *Person A* and *Person B* that after the procedure has taken place, one person will receive a reward of $100,000 while the other will be tortured. Who receives the reward, or the punishment, is a decision left for *Person A* and *Person B* to make. Locke would likely argue that *Person A* has survived in *Person B's* body and vice versa as we, along with our personal identity, follow where one's consciousness lies. For this reason, Locke would suggest that if *Person A* were to make the decision, they should request that *body B* receives the reward, as this is where *Person A* would survive, according to Locke. However, as Williams asserts that one's personal identity is located with continuity of the body, he considers that both persons remain in their own respective bodies but have simply undergone an extreme change in their psychologies. Williams would conclude that *Person A* should request that *body A* receive the reward and that *Person B* would similarly want *body B* to receive the reward.

Interestingly, Williams goes on to argue that this presentation (Locke's view) misleads our intuitions into suggesting that we go where our psychologies go, and that it instead *can be rational* to see the situation of a psychology swap.

Williams's version of the Mad Scientist (*Person A* survives in *body A* and *Person B* survives in *body B*) faces a significant challenge which raises doubt over the proposed physical criterion of personal identity. Suggesting that *Person A* survives in their own body contradicts our natural intuition of who we think and expect will wake up in *body A.* When presented with the first version of Locke's revised thought experiment offered by Williams, which is the case where *Person A* has the entirety of their consciousness scanned and copied into *Person B's* body who had moments before had all the contents of their consciousness erased, it is argued that our intuition has us naturally believing that *Person A* will survive in *body B.* Similarly, in Williams's 'complete' consciousness swap, when asked who will survive in each respective body, our intuition again has us ready to agree that *Person A* survives in *body B,* and that *Person B* wakes up in *body A.* Williams (1970) on the other hand, claims that when the scenario is not presented misleadingly our intuition leads us to a different conclusion.

However, Williams attempts to show how our intuitions can be misled into suggesting that we travel where our psychology goes. Williams is not saying that our intuitions always go the other way, only that we do not have to go with this intuition, meaning that it can in fact be rational to consider this scenario (the first presentation) of a psychology swap as simply a person going crazy and thinking that they are someone else. A point that significantly counts against the psychological view is that we could use the same equipment to create a scenario where we have *Person A's* psychology in both A's and B's body. If this were the case, the psychological view would be in trouble, because it does not match the logic of identity (as two of the same things would therefore be in different places at the same time). It is here where we see psychological theorists attempt to rescue their respective criteria, an example being Parfit focusing on logic and uniqueness, which was discussed earlier.

The charge that Williams faces here is that his claim (about which body he argues our intuition each person will survive in) is simply inconsistent with our actual intuitive assumption of who will survive in which body. Something seems amiss when a physical theorist suggests that *Person B* has survived in their own body and should therefore receive the money, regardless of the extensive change in their psychology that they have undergone. Rather, it presents a case in which we imagine someone swapping bodies. A step further, it seems absurd to award the money to *body B,* who physical continuity theorists claim is *Person B,* when the person displays a set of thoughts, memories, and so on that are exactly identical to what we saw *Person A* exhibit before the swap. According to the physical criterion, *Person B* (which physical continuity theorists consider as having survived in *body* B) should receive the money regardless of possessing a completely different consciousness. For most, this goes entirely against what our intuition says. Williams asserting that our intuition points to each person surviving in their original respective bodies is simply not a fair or accurate claim and holds no ground as our intuition in fact leads us to an opposing conclusion as to who survives in which body. Thus, arguing that our intuition supports the assumption that each person will survive in their original bodies does not provide sufficient support for Williams's criterion of bodily continuity. This leaves Williams with very little, if any support at all for his physical criterion up to now. Williams's argument that it can be rational to see *Person* A staying in his body despite getting a new psychology does not combat our actual intuitive response to the body-swap version of the thought experiment. However, it is interesting to note that Williams may well be right that there are other thought experiments to which we do not know how to respond (like that in which we end up with two psychological A's), but that does not affect the fact that we *do* know how to respond to this one, and in such a clear way that it provided insight into how our concept actually works. It is this that does significant damage to the physical criterion, even if it raises a problem for the psychological view (which Parfit attempts to solve).

Williams (1970: 161) acknowledges that there are details that expose the limitations of the consciousness swap thought experiment; however, he goes on to insist that if you want to present a

convincing body-swap thought experiment (although something which he does not want to do), then you should discount these details (such as *Person A* and *Person B* differing significantly). The concern and criticism here is that we should not, as Williams suggests, discount the case where *Person A* and *Person B* differ significantly from one another as this would be conveniently avoiding a crucial challenge presented against the criterion of bodily continuity. In an attempt to give psychological theorists a sympathetic hearing, Williams tries to describe how they make their case convincing, although proceeds to later argue against psychological theorists. However, discounting these types of details (dissimilarity between those in the scenario) does not do a physical criterion any justice, and instead casts doubt on a physical continuity view. It should be examined how the criterion of physical continuity stands against the scenario where both persons are significantly dissimilar.

Let us consider specific details added to the complete consciousness swap to demonstrate why a physical criterion cannot survive Williams's suggestion of discounting details such as the potential dissimilarities between the two persons in the scenario. Suppose *Person A* is a young female, a devoted mother, and enjoys ballet and painting. *Person B* is a male, a competitive bodybuilder, and spends his free time watching movies. After the procedure has taken place and the consciousness of both persons has been swapped with one another, we would see *body A* display features of the male bodybuilder, and *body B* would appear to be the devoted mother who enjoys ballet. To emphasize, *body A* would appear to be a female who is a competitive bodybuilder, and *body B,* a male and muscular body, would appear to be a young mother who is passionate about ballet. Ignoring these stark differences would prove difficult, if at all even possible. For many, intuition would dictate that *Person A* and *Person B* have swapped bodies and survived in the body of their counterpart. Additionally, to sufficiently defend a physical criterion, one would be required to argue and explain *why* we ought to agree that each person survives in their respective bodies, especially when the conclusion contradicts with our natural intuitions. A traditional physical criterion's inability to provide a reasonable explanation as to why we should abandon our intuition and agree that both

persons survive in their respective bodies is problematic and leads us to question the validity and strength of the proposition that personal identity lies with bodily continuity.

Here I would like to return to detail a notable concern with Williams's physical continuity view. As mentioned a little earlier, Williams asks us to ignore the case where *Person A* and *Person B* are remarkably different (with regards to their mental content) and instead requests us to consider the two persons in his thought experiment as relatively similar to one another. The charge here is that Williams appears to be relying on the very element he attempts to refute. Williams is relying on the fact that the two persons' psychologies (and mental content) in *The Mad Scientist* are relatively similar, presumably because this will aid in leading us to consider that a psychology swap has occurred and *not* a body swap. Put more clearly, there would be less trouble in accepting that *Person A* survives in *body A* but with a slightly different psychology compared to accepting that *Person A* survives in *body A* but with a completely different psychology, in fact that of another person.

While a thought experimenter can choose which features to include or exclude, a thought experiment should only be considered if by choosing which features to include or exclude, it is not done in a manner that hides some deep impossibility. For Williams' physical criterion to hold any merit, it ought to withstand any detail of the psychological content of the persons involved in the thought experiment. It is clear that Williams fails in this regard. As he admits, we should ignore the possibility of there existing significant differences between the psychologies of *Person A* and *Person B.* However, this demonstrates the Williams himself is depending on elements of the psychological view as a way to enable his physical view to work. This illuminates serious trouble for Williams and how he is unable to provide a thought experiment that can independently defend the bodily criterion, without requiring any role of a psychological view.

It is also worth noting that the success of Williams's thought experiment seems to rest on his need for the guarantee that *body A* now possesses memories (and other mental content) that truly belong to *Person B* and are not instead artificially manufactured memories, beliefs, and so on, that

were merely manufactured to match *Person B's* earlier memories, beliefs, etc. (Williams, 1970: 162, as cited in Beck, 2014). This suggests that Williams agrees that the memories that were swapped from *body B* to *body A* do in fact belong to *Person B*. It then follows that if these memories belong to *Person B*, then once the memories are placed into *body A,* Williams is forced to admit that *body A* now possesses *Person B's* memories and other mental content, that it is not still *Person A* with an altered psychology. Williams would be unable to argue that *body A* now is still *Person A* who now possesses different memories and mental content but that those different memories and other mental content still belong to *Person A.* As Williams has assigned the concept of ownership – the memories and other psychological content of *body B* belonging to *Person B* – Williams unintendedly corners himself into saying that the memories, beliefs, and other mental content in *body A* belong to *Person B,* and *not Person A* as he would like to have us believe.

When assessing how Williams fares against the four features of identity, and to provide a fair and holistic discussion, it is worth acknowledging that there are cases where the feature of self-interested concern fits well with Williams's bodily criterion. This has been discussed in detail above, so a brief synopsis will be offered here. In a scenario like the second presentation where a person is told they will be tortured, but will undergo complete amnesia moments before, the person maintains a sense of concern for themselves in the way that they are concerned about what will happen to their body, believing that it will still be *them* who is tortured. In this scenario, even extreme changes to their psychology does not ease the feeling of dread or fear, and the person still feels a sense of (self-interested) concern for what will happen to their body in the future in the way that they feel *they* will still be the one experiencing the future pain/pleasure. Irrespective of whether I am told I will not remember being tortured or that there will be significantly altered mental content in place of my current psychology, I still intuitively fear the imminent torture and consider this as that *I* will be the one undergoing the torture. At this point, Williams's criterion seems to align neatly with this feature of personal identity, as well as our intuition.

However, unfortunately for Williams, his physical criterion is in fact not immune to all challenges with regards to aligning with the feature of self-interested concern. There is a concern that our intuition at times has us more concerned about where our psychology goes compared to our body. As an example, Williams's first presentation demonstrates this intuition where if *Person A* and *Person B* were told that they would have their psychologies switched and need to decide who gets tortured and who receives a large sum of money, our intuition pulls us into believing that if we want the reward (and subsequently avoid the torture), we ought to follow our psychology, meaning that *Person A* would opt for body B to receive the reward after the procedure. A scenario like this demonstrates a stronger concern for our psychology compared to our body, meaning that our self-interested concern also follows our psychology, a conclusion which is problematic for Williams where it seems that our physical body plays no role (at least in a scenario as above) when contemplating what we think will happen to *us* in the future. If we were to follow our intuition and reward *body B* with the belief that we are rewarding *Person A,* Williams would be in trouble of seemingly depending on a *psychological* criterion of identity instead of a physical criterion.

Moving on, the features of moral responsibility and compensation will be mentioned in conjunction. As we saw above with Locke, regardless of whether one remembers committing an action, we naturally feel inclined to hold them responsible for the act they did in fact commit. Similarly, we typically think it is only fair to reward someone for their good deeds, irrespective of whether they actively remember performing this action. This seems to work in Williams's favour.

However, consider the case of a swap between *Person A* and *Person B's* psychology. When *Body A* wakes up with the psychology of *Person B* and possesses all the beliefs, desires, intentions, and so on, and goes on to commit a murder, most of us would argue that it is *Person B* who has committed the murder, while simply occupying the body of *Person A*. Similarly, when we witness *body B* donating to charity, an action commonly performed by *Person A* (prior to the procedure), we would look at this as *Person A* doing this good deed while existing in *body B.* In terms of the murder, we

would be inclined to hold responsible *Person B*, who just happens to be in *body A.* If *Person B* were in their own body, or in *body A*, we would still be inclined to hold responsible the person of the psychology who committed the murder, not simply the physical body that was involved.

Taking this one step further, suppose a procedure is performed that transfers all of *Person A's* psychology into *body B*, and vice versa. Then, *body A* (with *Person B's* psychology) goes ahead and commits a murder. Shortly afterwards, a second procedure transfers the psychologies again, resulting in *Person A* existing in *body A,* and *Person B* in their own. You are now asked who should be held responsible for the murder. Williams would argue that *Person A* should be held accountable. However, *Person A* has no memory of this act, whereas *Person B* does. In fact, *Person A* watched *Person B* commit the murder in their (*Person A's*) body. It seems nonsensical to hold *Person A* accountable simply because it was *body A* that was involved in the crime. We are intuitively drawn into wanting to hold *Person B* responsible, as this is where the intention to commit the murder, as well as the memory of this act are seated. Here we see that Williams's criterion is unable to meet these features of identity without encountering difficulties, and to a degree fails to provide a criterion of identity that does what a theory of identity ought to do. For Williams's view to survive, he would be obligated to accept a psychological criterion as necessary for continued personal identity.

Eric Olson and the brainstem

Consider the following: a person has lost all their mental content and mental capabilities and are now in a vegetative state. While the physical body may still be functioning independently (without the need for life support), there is no sense of agency or surviving (psychological) characteristics of the person. The person will shortly undergo a lifesaving procedure whereby they will receive a healthy and functional cerebrum from a recently deceased person, the donor, while each respective brainstem will remain in their original bodies. The operation is a success, and the person wakes up in

their original body but with the donor brain. Traditional body theorists are confronted with a scenario where one physical body contains parts of two different bodies. Logically, the body theory cannot suggest that the persons of both bodies have survived and continue to exist as this would be nonsensical. We are forced to accept that one person exists but are tasked with determining who that person is.

A fellow opponent of the psychological continuity theory, Eric Olson (1997a) denies the need for a person's full body, or even most of their body be required for continued personal identity. Olson (1997a) acknowledges that our intuition has us believe that a person would go with their transplanted brain, implying that the person who wakes up after the procedure is the donor body who had their brain transplanted into the person who was originally in a vegetative state. However, Olson (1997a) argues that our intuition is in fact false and misdirected, and that believing that the person who survived is the donor person because their intact brain was transplanted is incorrectly based on moral and practical considerations instead of metaphysical concerns. While also suggesting a form of physical criterion, Olson (1997a) considers the seat of personal identity, the necessary criterion, to lie specifically with the brainstem and not the whole brain itself. According to Olson (1997b), a person continues to exist if they possess their functioning brainstem as this is the part of the body that is responsible for maintaining essential functioning of us as human organisms such as breathing and our hearts beating. In Olson's favour, this view seems to provide a neat and satisfactory response to the scenario where a person is being kept alive through artificial life-support. Recall the above discussion regarding our intuition in cases like this. For many, if not most of us, our loved one has passed on and no longer exists in the mortal world. Their body is simply being kept alive mechanically, but they as a *person* have not survived. Olson would confirm that the person has indeed not survived as although they are still in possession of their brainstem, their brainstem is not functioning in a way as to sustain essential bodily functions independently. In this way, Olson can provide neat and plausible explanation for personal identity and why we consider the person in a vegetative state to have no longer survived. Olson would agree that in cases like

these, moral considerations of the person fall away, although interestingly, Olson does not think *anything* (=any *thing*) has *not* survived. It is here where we start to see a change in focus where although Olson calls his Biological View a theory of personal identity, this is in fact not what it is. Rather, his Biological View is a theory of *our* identity, and we are not persons according to him. Olson appears more focused on detailing the biological aspects of our identity as human beings, rather than discovering the essence of the continued identity of a *person* that is distinct from other persons.

It is not unreasonable to accuse Olson of making the same mistake that has been presented to challenge Williams, namely that his physical criterion undeniably works against our natural intuition of personal identity. Imagine a case where *Person A* has their entire consciousness and all mental content erased and then replaced with *Person B's* consciousness and mental content. Olson is forced to claim that as long as *Person A* is in possession of their own (and functioning) brainstem, it will be *Person A* who survives, regardless of the total change experienced in their consciousness and mental content. It is troubling to believe and agree with the conclusion that *Person A* survives due to the simple fact that the surviving brainstem is from their body.

Turning back to the scenario where a person receives a donor brain, Olson's theory hangs on a very technical detail – the brainstem. The identity of the person who wakes up after the procedure depends on which person's brainstem is present in the surviving body. Again, assuming that the procedure is successful, and the memories and other mental content of the donor brain remain intact, regardless of which brainstem is in the surviving body, whoever wakes up will have the memories and psychological features associated with the donor person. I will detail this further for clarity. On the one hand, the person who wakes up will either be the original vegetative person but with the donor's mental content. On the other hand, the surviving person will be the donor person but in the original vegetative person's body. Both will exhibit the same mental content such as memories, thoughts, desires, and so on. It seems absurd that the identity of the surviving person

hinges on the ownership of the brainstem. Without being aware of the ownership of the brainstem (whether the brainstem is of the original body or the donor body), intuition would still have us convinced that the surviving person who wakes up is the same person to whom the donor brain belonged to, and that the person who was in a vegetative state and received the brain has not survived. In the case where both possible surviving persons appear to be the same as one another, it is therefore reasonable to argue that the brainstem on its own is insufficient for personal identity, and that there must be something apart from the brainstem that is responsible for surviving personal identity.

One final scenario will be detailed to demonstrate the shortfall of Olson's brainstem criterion. For the purpose of this thought experiment, let us consider that *Person A* was in an accident that led to irreparable damage to their brainstem. Fate has it that a match has been found for *Person A* and they will soon receive a new, functioning brainstem from a donor, *Person B,* who recently suffered from a fatal heart attack and has been pronounced dead, although their brainstem was miraculously not affected or damaged. Surgeons successfully salvage *Person B's* brainstem, healthy, intact, and functioning, and transplant it into *Person A's* body. By arguing that the brainstem is responsible for continued personal identity, Olson is forced to admit that the person who wakes up from the procedure with the same continuing consciousness as *Person A,* remembers being *Person A* before the operation and beyond, is instead *Person B.* Put differently, Olson is forced to admit that *Person B*, although having been pronounced dead before the procedure just a few hours before, is the person who wakes up and survives the operation, and that *Person A* has not survived. The implication of this is that the surviving person possesses all the body of *Person A* except for the brainstem from *Person B,* as well as the entire consciousness and all mental contents of *Person A.* Rationally, we can see that the surviving person has significantly more elements from *Person A* than *Person B.* Not only is Olson's claim of *Person B* surviving being in great contention with our intuition that tell us that it is *Person A* who has survived and received a brainstem transplant from a donor

body (*Person B*), it also seems nonsensical and even far-fetched to suggest that *Person B* survives instead of *Person A.*

Olson allows for abortion even though the "person" seemingly exists. He avoids the criticism of killing a person by drawing a distinction between the questions of "What am I" and "What matters" and argues that what matters might be what matters to the person (but as a fetus, you are not yet a person). As a fetus, you are a "pre-person", not yet a full person. A fetus is conscious, but not self-conscious. So, an abortion would be killing a pre-person, not a person. What you are killing is not what matters – it's not a person. A traditional animalist will have a problem with abortion because they will see it as killing a person. However, Olson avoids this by drawing the distinction between the two questions. Olson says that personhood (a human possessing moral status) is just a phase that a human goes through – *person* is not a substance concept at all, so there is no such substance (= thing) as a person, given his metaphysics. Rather, *person* is a functional concept, and substances are meant to be fundamental things. Olson's argument here echoes what we saw earlier with Locke, whereby Locke believes human beings and persons to be different kinds of things, which Locke believes avoids the charge of two things (the same things) existing in the same place at the same time, which is logically impossible. We witnessed Locke explaining that a human being is one thing (and potentially a combination of two things, namely a soul and a body), and that personal identity is a separate thing that is neither the human body, nor is it a matter of either soul identity or body identity. However, we are not provided with a clear explanation as to what this third thing ("the person") is. We know what it is *not,* but clarification as to what it *is* has been omitted. The concern which follows is that we now have too many elements involved in the puzzle of personal identity where a holistic conception of a person (beyond their identity as simply a human being) appears to consist of several pieces, which seems to complicate what it is exactly personal identity is and suggest that by 'splitting' these different entities, we have, for now, lost sight of what exactly is responsible for personal identity.

An important challenge that is worth raising against both Olson and Locke would be the question of when a *person* (Olson's person phase) comes into existence. At what point does a fetus, or an entity at the 'pre-person' phase become a full *person*? Does the human organism become a conscious person at a certain age, or when they have developed a specific capacity of functioning? If so, what would this age (or capacity) be, and *why* this specific age (capacity)? Suppose the answer to questions like these is that once a human has developed the capacity to speak using sentences with five or more words, a conscious child acquires *person* status and can now be held morally accountable for their actions. The first issue with this suggestion is that this timestamp seems somewhat arbitrary. Comparing a little girl a day before she developed the ability to create five-worded sentences and a little boy who has this reached this developmental milestone, the two seem identical (to some degree) in terms of their overall development and maturity. It seems odd to suggest that if both children hurt one another, we can only hold the little boy accountable because he is *technically* of the developmental capacity where we can hold him accountable (simply because he is able to create slightly longer sentences than the little girl), but the same cannot be said for the little girl who currently speaks using sentences consisting of four words. Olson attempts to defend himself by claiming that this is merely a moral issue, instead of a metaphysical issue. It may be the case that there is a moral way of solving this issue, even if there is no metaphysical solution. However, a metaphysical solution would be preferable, which Olson's theory fails to offer while most of the other theories work to offer a metaphysical solution to issues like these. This simple example works to demonstrate the trouble Olson (and Locke) would face when trying to specify when exactly a human qualifies as both a human being and a *person.*

Let us investigate how well Olson's criterion satisfies the four features of identity. Looking first at the survival feature, if a person was to no longer possess their brainstem, they would simply not survive. It would not be the case that one's *personal identity* ceases to exist, but the human being themselves would not survive as the brainstem is imperative for functioning survival on its own. Now, in the case where a person experiences severe damage to their brainstem (as a result of injury

or disease, for instance) but undergoes a successful brainstem transplant while maintaining their psychology, we would consider the person as having survived. Olson's criterion works against our intuition by arguing that the person we once knew would not have survived. A simple case like this highlights the trouble Olson's criterion encounters and shows that it ultimately fails to satisfy the feature of survival.

Olson's criterion also falls short with regards to the features of moral responsibility and compensation. If a person commits a murder and between the time of the crime and their date in court they undergo a brainstem transplant, Olson's criterion would have no metaphysical answer as to why they were held responsible. Olson would have to say that it makes practical sense even if there is no deeper reason for it. However, we would still want to hold the person responsible for committing the murder, regardless of the brainstem transplant. In fact, we would naturally not take the brainstem transplant into account when working on assigning responsibility. Intuitively, it is irrelevant whether the brainstem originally belonged to me or a donor. Consider another example where a person dies of a heart attack, but their brainstem is salvageable and transplanted into another person whose brainstem is deteriorating from disease. The patient, with their original psychology, survives the operation and goes on to find the cure for cancer five years later. It seems absurd to claim the original brainstem donor as the person who cured cancer. We would undoubtedly consider the patient (who received the healthy brainstem) as the person who deserves the credit. These two examples demonstrate that who the brainstem belongs to is irrelevant for personal identity.

Finally, we can see that Olson's criterion also encounters trouble with the fourth feature of identity – self-interested concern. Suppose my friend and I are about to undergo a procedure where we would be exchanging brainstems (but retain our own psychologies). Typically, we would not see this procedure as a body-swap, rather just a brainstem-swap. We would expect to each survive in our respective bodies, simply with a different brainstem. It would not be the case that while

preparing for the surgery, I am thinking about what my life would be like when I wake up in my friend's body. I would be thinking about what will happen to *my body* in the recovery period. I will be concerned about how *I*, in my body, will recover from swapping my brainstem for my friend's. Despite caring for my friend, I will feel a special concern for my own recovery (in my own body) more than their recovery. It is this special self-interested concern that shows that even in a case we would intuitively consider a brainstem-swap, we believe we will remain in our own bodies and that our personal identity will not follow our physical brainstems, demonstrating that our brainstem is an insufficient criterion for personal identity as it does not satisfy the feature of self-interested concern.

Marya Schechtman and the Social Element

In *The Constitution of* Selves (1996) and her later work, *Staying Alive* (2014), Marya Schechtman proposes two versions of her view – an earlier and a later one. Schechtman's earlier view was a narrative approach, whereas her later theory, or her cluster approach, only has narrative aspects. Let us examine both approaches, beginning with her (earlier) narrative view. Schechtman's (1996) Narrative Self-Constitution View explains that a person's identity is created by them establishing an autobiography (narrative) of their lives. Schechtman (1996: 94) explains her view as follows:

> *…the difference between persons and other individuals...lies in how they organize their experience, and hence their lives. At the core of this view is the assertion that individuals constitute themselves as persons by coming to think of themselves as persisting subjects who have had experience in the past and will continue to have experience in the future, taking certain experiences as theirs. Some, but not all, individuals weave stories of their lives, and it is their doing so which makes them persons.*

A person survives as one and the same *person* over time if they possess a self-understanding, or a sense of self, where they have the capacity to understand experiences, thoughts, and actions as part

of the story (narrative) of their lives. This would mean that each thought, experience, and so on of one's life "fits" into the story of their life in a coherent and meaningful way and "seeing experiences and actions as part of an intelligible whole" of the person's life where "actions and experiences are yours in so far as they fit meaningfully into your life story" (Beck, 2016: 118).

Schechtman's (2014: 110) later version of her view is her Person Life View (PLV) which can be detailed as such:

> *To be a person is to live a 'person life'; persons are individuated by individuating person lives; and the duration of a single person is determined by the duration of a single person life.*

Schechtman's (2010; 2014) PLV shifts emphasis from the person's own narrative to the narrative of those around them – to their 'place in person-space'. Schechtman's earlier view followed the typical development an individual experiences from an infant-stage through their lives and accepts that "humans can live very different sorts of lives, but points to a very general shared form of development" across cultures (Beck, 2015: 8). The PLV differs from Schechtman's original narrative view whereby "persons are defined in terms of the characteristic lives they lead" (Wagner, 2015: 141), and that "persons are individuated by individuating person lives; and the duration of a single person is determined by the duration of a single person life" (Schechtman, 2014: 110). This view emphasises that a person life is lived "in a culture and in interaction with other persons", and a vital part of what it means to be a person is to interact and be engaged in meaningful ways with other persons around oneself (2010: 279; 2014: 114). While this social aspect has been mentioned in her earlier views, Schechtman places this element as the forefront of her PLV. Schechtman stresses the importance of not only a person being aware of themselves continuing to be the same person, but that it is also critical for people external to you (those in society) to recognize and acknowledge you as being the same person over time (Beck, 2016: 128).

Schechtman goes on to describe the concept of the 'person-space', which Wagner describes as "being able to attend a room of shared attention and meaning" (2015: 141). Schechtman (2014) proposes that the concept of a person life should be viewed as a cluster concept that incorporates biological, psychological, and social elements which function together. Where Schechtman differs from physical and psychological theorists as seen above is that in this cluster approach, none of the three criteria are sufficient on their own for sustaining a person life (Beck & Oyowe, 2018). Instead, the three features typically work together in a way that supports continuation of a person life, although Schechtman (2014) allows for these elements to come apart where a person life is preserved even in the absence of any one of the three aspects.

A cluster theory with this social criterion has a person's sense of self play a crucial and central role in their personal identity and is categorized as a person being able to describe themselves (and the story of the life) in a way that they would when they meet new people, for instance. Moreover, Schechtman's view goes beyond just an intimate sense of oneself. This approach (Schechtman, 2014) differs from her colleagues where instead of a rigid and all-or-nothing concept of personal identity, she explains personal identity as a fluid and flexible concept that constantly develops along with the experiences of one's life. Where continuity theorists such as Locke, Parfit, and Williams require sameness of a criterion which in the end proves rather restrictive and simply unrealistic (such as requiring the same body, enough of the same memories, and so on), Schechtman allows space for the inevitable and inescapable changes we experience to our body and psyche throughout the span of our lives while maintaining the same persisting personal identity.

Schechtman's view is in contention with previously discussed bodily and psychological continuity theories where she argues that neither alone are sufficient for personal identity. Schectman accuses previous theorists of omitting an essential part of identity. The missing aspect Schechtman emphasizes is a social dimension which refers to a person's capacity of engaging and interacting with others within a social dynamic (1996: 133). For Schechtman (1996), it is not only crucial for a person

to recognize *themselves* as being the same person, but that others in society also recognize and consider you to be the same, continuous person. This is not to say that Schechtman entirely rejects the importance of the role of bodily and psychological continuity. Rather, she (1996) denies that personal identity can be attributed to these criteria alone. It is then also worth clarifying that Schechtman is not arguing for a social criterion in isolation.

To work around the challenges and limitations the theories discussed above encounter, Schechtman attempts to offer a view that is somewhat less restrictive. Schechtman (2014) explains that not all three criteria – biological continuity, psychology continuity, and the social criterion – are required to maintain personal identity. She states that that is the normal state of affairs for there to be all three continuities, and that is how a person life is to be understood. Each of the three functions is an abstraction from the usual cluster. Schechtman (2014) accepts that possessing two of the three criteria can be sufficient for survival and continued personal identity (put differently, there can be survival without one of the three criteria being fulfilled, provided that the other two criteria are present). For clarity, this understanding suggests and argues that a person remains the same person as their earlier self if they, for instance, possess the same consciousness (memories and other mental content) and society recognizes them as the same person. Without any hesitation, this example seems to align neatly with our intuition and our expectation of who we think survives. Let us consider this case more closely to examine the affect bodily continuity may have on the identity of who we consider to have survived. Suppose the person maintains continuity of their consciousness, possesses all the same memories (within realistic limitations), is recognized by society as the same person, but they woke up that morning in a different body. Thought experiments similar to this highlight our intuitive response to believe that a body swap has taken place where the person (to whom the memories, beliefs, and so on, belong) has survived, simply in a different body. Rightfully so, if someone woke up one morning and told us that they were our best friend, thought and spoke like our best friend, has the same memories we expect our best friend to have and can recall significant past events when interrogated, we are typically inclined to believe

that unless the person has simply gone crazy (as Williams suggests), it would appear that something along the lines of a cerebrum-transplant has taken place and that our best friend, while still the best friend we recognise, has woken up in a different body. This is just one illustration that demonstrates that Schechtman may be on the right track with her combination and narrative view.

Another example of a case that works in Schechtman's favour is the common inevitability for many of us – dementia, Alzheimer's Disease, or other neurological memory diseases. Consider your grandparent who is well into her 70s has been living with symptoms of severe dementia for the past year and is subsequently losing more memories of her life each day. Although your grandparent no longer possesses a meaningful degree of psychological continuity, the criterion of bodily continuity is met. Usually when a person suffers from memory loss, whether a temporary occurrence or due to a progressive neurodegenerative disease, we still consider the person as existing along with us, just with memory difficulties and complications. This way, two of the three criteria – bodily continuity and social recognition – are fulfilled and therefore, according to Schechtman's view, satisfy the necessary criteria for persisting personal identity. As with the case examined just moments ago, this scenario is also consistent with the way we typically and intuitively view personal identity. The alignment between us believing that the person the grandparent is continues to exist and Schechtman's theoretical answer that the grandparent retains their personal identity, irrespective of the tremendous changes and loss to their memory, provides a level of support for her cluster view. Importantly, that same support does not work for her earlier narrative view as the demented individual would no longer have an autobiographical narrative, and so would not qualify as the same person, and this is what Schechtman is trying to amend with her PLV.

A critical concern, however, with Schechtman's view, is the degree of influence that is given to the criterion of social recognition and whether society views a person as the same person who continues to exist over time. In other words, there is a concern that society could perhaps make a mistake regarding one's identity such as no longer treating them as the same person when there is

some factor that makes them intuitively still the same or treats them as the same when they are not. Kristie Miller's 'Eagle people' (2013: 95-96) neatly demonstrates this concern:

> *Suppose, through the wonders of science fiction, I am transported to a world in which there are humanoid entities who seem much like myself. At a certain age each humanoid is killed by being beheaded, but each expects to survive this process by becoming, at that time, an entity that is very much like what we would call an eagle. The community has a range of person-directed practices such that they each anticipate the experiences of some particular eaglish entity, and where the society recognises that eaglish entity as being the same person as the humanoid that was killed. After the beheading, for instance, the eaglish entity attains rights and responsibilities it did not previously have. When I arrive in this community, I am not clear what to make of this strange procedure. I am a strong conventionalist, yet I find it intuitively implausible that these humanoids survive the procedure as eaglish entities. What I am sure of, however, is that were my head to be removed, I would not survive the procedure, and this worries me since I am approaching the age of beheading. Soon after, I come to discover that there is a herb I can ingest that will modify my psychology so that I will come to anticipate the experiences of some future eaglish entity, have self-concern towards that entity, and so forth. If I am fairly certain that I will shortly be beheaded, ought I to take the herb, on the understanding that if I do I will survive having my body killed?*

Suppose I choose to ingest the herb in the near future. Regardless of the changes that I am aware will be made to my psychology, my intuition continues to pull me into believing that I will not survive the beheading, especially not as an eaglish entity. However, society firmly believes that after ingesting the herb and being beheaded, I will survive as an eaglish entity, and will treat this eaglish entity as *me.* It seems absurd to suggest that this eaglish entity, whom I would not consider to be *me,* not only be treated as *me,* but that it is understood that this eaglish entity will be continuous with *me.* The extreme counterintuitive response I have of being told that I will survive as an eaglish

entity raises trouble for the allowing the criterion of social recognition to be given such importance in determining the continuation of one's personal identity. Furthermore, society would need a reason to base their belief that it is in fact *me* who survived as the eaglish entity as it cannot be the case that I am believed to survive as the eaglish entity simply because society thinks so. Suppose that the eaglish entity possesses a sufficient degree of psychological continuity to the humanoid I was originally before the beheading. The social criterion would appear to function as the deciding factor in whether I have survived as the eaglish entity or not. This would again seem absurd if the eaglish entity is convinced that they were the original humanoid but, for an unknown reason, society thinks that the eaglish entity and original humanoid do not share the same personal identity (suppose they believe the psychological content of the eaglish entity to have been artificially manufactured during the beheading process). To add to the concern about the level of importance given to the social criterion, it is reasonable for me to have trouble with society treating the eaglish entity as *me* where I firmly believe that I will not survive the beheading at all, let alone as the eaglish entity. I would intuitively be searching for the reason that society believes that the eaglish entity is *me,* therefore placing my belief of where personal identity rests on a criterion other than the social element.

This presents a problem for Schechtman that challenges at least either her set of three criteria, or the importance given to the social criterion. There is a sense of conventionalism here that Schechtman will not want – that society's whims may determine survival, as it seems intuitively correct that societies can in fact be mistaken. It seems extremely troubling to suggest that a person's identity can rest in the hands of society who, at their own will and discretion, may choose to 'decide' whether you are still *you*, and that even when society makes this decision with the best intention, their conclusion of whether you are still *you* could in fact be wrong. An example to demonstrate this would be to imagine a case where a person (person A) was involved in a car accident and although is physically alive, they no longer have any psychological functioning. Luckily, a device exists that can extract the person's mental content and transfer it into the brain of another body (body B, where

the person of this body is also braindead). The doctors scan person A's mental content and transfer it into body B. Shortly after, the family of person A comes to see their loved one but is not told about the psychology transfer. Asked if they think their loved one has survived, many may argue no. Where the family do not recognise person A as surviving, person A would be considered as no longer existing, and this is consistent with Schechtman's cluster theory where an insufficient amount of the three criteria have not been met for personal persistence (as only one criterion has been met – bodily continuity). However, if (although counterintuitively) the family sees person A as having survived albeit in a different, vegetative state, Schechtman's cluster theory would be forced to conclude that person A survives as two of the three cluster criteria have been satisfied. The trouble with this is that for many, this is deeply counterintuitive. Furthermore, it should be mentioned that although some might argue that giving person A's family the final say (rather than strangers in society) in determining person A's survival is justified as they have a deeper understanding or connection to the person, it is not necessarily the case that all members within a family will be in agreement, which again seems to suggest that society is in a position to 'choose' whether person A has survived, regardless of whether they are closely connected to who person A is (or was), or complete strangers.

Now, consider the family learns about the transfer. They are told that person A's body is in a vegetative state, and that person A's psychology and entire mental content has been successfully transferred into body B, and that the person who has awoken in body B has person A's memories, personalities, beliefs, and so on. The family speaks to the person of body B and come to find they are exactly the person they knew person A to be. Here the psychological criterion is met, so if the family recognises body B as person A, person A therefore survives, just in body B. However, if the family reject that this is person A, then, according to Schechtman's cluster view, body B does not survive as person A (as at least two criteria are not met – the physical and social criteria), but rather as a person who is psychologically the same as person A. This is deeply questionable as it works against our intuition where we could be having a conversation with the person of body B, but because

society does not consider them as person A, we are forced to accept that we are not having a conversation with person A, although we are discussing shared memories and experiences we had with the person of person A. This example works to illuminate noteworthy challenges Schechtman's cluster view faces.

A final example of the trouble Schechtman faces that puts pressure on her social criterion is whether the family views both person A and body B as person A. In other words, suppose the family views the vegetative state of person A's body as person A surviving, while *also* seeing person A as surviving in body B. As the identity in these two cases hinges on the social element, this would be a technical possibility, although deeply counterintuitive and illogical. This highlights a significant concern that challenges Schechtman's social criterion. In these cases, it appears that what matters for personal persistence cannot have the social criterion as the deciding factor. This points to something else being responsible for personal identity and that we cannot put as much emphasis on the (external version of the) social criterion as Schechtman suggests.

Potential Resolution

Thus far, philosophers are yet to arrive at a conclusive argument for where personal identity lies. For some, the question has shifted away from determining which criterion it is that is responsible for persisting personal identity, but rather what it means to *survive,* and what it is that matters for *survival.* When a person experiences significant changes, such as a fundamental change in their religion, we may speak in a way to say that the person as we knew them has changed, but we do not literally mean that they are a different *thing.* Philosophers often take for granted that survival requires (sameness of) personal identity. However, the direction of this discussion will investigate the possibility that personal identity is not required for survival, to mean that I can survive without the continuation of sameness of personal identity.

When comparing Schechtman's narrative view and her PLV, it can be argued that her earlier approach captures the four features better, as well as avoiding other concerns the cluster view faces. Here I will argue that her focus on someone's own narrative is more important in achieving what a theory of personal identity is meant to achieve and that it is foregrounding the social element and making it the most important and dominant feature that leads Schechtman into trouble. I agree that Schechtman's cluster concept approach carries validity, but the concern is that Schechtman places too much emphasis on social continuity, where it can be argued that psychological continuity (the kind we witnessed in our discussions of Locke and Parfit) is more important and plays a more fundamental role (although social continuity is still relevant, of course). Additionally, Schechtman's earlier narrative theories can be viewed as being closer to a psychological account than her later one. In order to create a narrative of one's life, a person is required to possess psychological continuity, or consciousness, making this element fundamentally vital in Schechtman's earlier view. This way, a person's self-concept possesses more importance than that of an external social element, as described in Schechtman's PLV.

Let us examine how Schechtman's narrative view captures the four features better than her cluster approach, with a particular focus on compensation and moral responsibility. Suppose it is the morning of my graduation and upon waking, I discover I have woken up in a body that looks nothing like mine (for the sake of the argument and clarity, let us refer to this new body as a male body). I have the exact same mental content as the night before when going to sleep. My parents are astonished when they see a man in the house claiming to be their daughter who is set to graduate. I know it is I, *Candice*, despite being in a body that is not mine. Schechtman's narrative view works better than her cluster approach here where, with enough time, it is reasonable to expect that my parents can be convinced into believing that this man sitting across them at the table is in fact me, *Candice*, sitting and speaking to them and that this would be achieved by detailing the narrative of my life. If I, in this male body, could provide a detailed and coherent story of my life, it is likely that my parents will realise that a bizarre body change had occurred overnight and that despite this, I,

Candice in a male body, will be graduating later that day and receiving the degree that I as *Candice* earned. We would not view the male body as an ordinary male student receiving a degree that is not *Candice's,* therefore satisfying the feature of compensation. On the other hand, Schectman's cluster approach would face difficulties in this scenario where although there is psychological continuity, my identity hangs on the decision of whether society recognizes this male body as me, *Candice*. As we have seen with other scenarios, it would be bizarre for me to know that I am me, but for society to tell me otherwise. Schechtman's cluster concept would also face trouble if some in society see the male body as *Candice,* and where some do not as we could, of course, not have both options be valid simultaneously.

Let us use the same scenario in relation to moral responsibility. I wake up as *Candice* but in a male body. Last week, I (*Candice* in my original body) received a parking ticket. As my psychological content is exactly intact and receiving the parking ticket is in line with the narrative of my life thus far, it is not unreasonable to hold me, *Candice* in the male body, responsible for paying for the parking ticket. I am able to recall the sequence of events that led to me receiving the parking ticket, as well provide a detailed and coherent synopsis of my life's story, we can argue that I simply woke up in a different body as a result of a bizarre occurrence. However, Schechtman's cluster theory casts doubt on whether the person in this male body ought to be held accountable for the parking ticket. There is psychological continuity in this case; however, the person in the male body will only be required to pay the parking ticket if society recognises the person as *Candice*, which I know to be true. If society tells me that I am not *Candice,* I would walk away feeling lucky to have gotten away with not needing to have paid *my* parking ticket, but also disturbed that I am not being recognized as *Candice* when I know that is in fact who I am. Society would also be in a position where no one is being held responsible for the parking ticket, which would be absurd. Ultimately, we see that Schechtman's narrative approach captures the features of compensation and moral responsibility compared to her cluster approach.

Moving on to the feature of survival. It cannot be contested that regardless of any degree of physical or psychological changes I may experience in my lifetime, if I am able to consciously argue that I as a *person* have survived, it would seem absurd to allow an external social element (as seen in the cluster concept) to challenge this and argue that I have in fact not survived. The absurdity of a scenario like this would be to imagine sitting in front of your friend after being involved in a car accident, having a conversation with them, but being told that you did not survive as you have lost various portions of your memories. It is also here where we see the importance of a psychological criterion creeping up in Schechtman's view, where (my) conscious recognition of my own survival seems to, and should, supersede an external social view as to whether I have survived or not. Schechtman's later view of how a person life fits in with those around them does not sufficiently capture the feature of survival in the same way that her narrative view does where if I am able to place memories, thoughts and past actions of mine in a way that fits meaningfully in the narrative (story) of my life, I would be justified in arguing that *I* have survived, which appears to fall more in line with how we naturally consider and speak about survival.

Let us discuss the last remaining feature, namely self-interested concern. If I were told that tomorrow I would be tortured, I would intuitively fear the impending pain. Regardless of any type of change to my physical or psychological attributes, I would nevertheless view the torture as about to happen to *me* and that *I* will be the person experiencing the pain. One step further, suppose I will have my entire psychology erased and those around me argue that I will not be the one who will be tortured (although it will be my body that will undergo the torture). In this case, we can consider two of the three aspects of Schechtman's cluster concept to be missing. However, this still does not take away the fear that I am feeling about the upcoming torture. Being told that I am not the person who will be tortured goes against my natural intuition about me being the one who is about to be tortured. The eagle people case, as discussed earlier, also works to support this concern. Even if it were the case that I ingest a herb that will alter my psychology in a way that has me believe I will survive as the eagle-like creature, and that this is what society believes too, from an external

perspective I cannot escape the intuitive concern of what will happen to me at the time of the beheading where I feel certain that *I* will not survive. Society's view that I will in fact survive simply does not match the self-interested rational concern I have with my survival from the beheading. Schechtman's later view is unable to suitably explain why regardless of whether I experience a significantly altered psychology, or have others tell me I am not the person who will be tortured, why I continue to feel a deep sense of impending fear. Rather, her earlier view fits more neatly with our intuition that I will be the person who will undergo the upcoming torture, despite any physical or psychological changes I may experience before or after the torture.

Finally, I would like to consider the argument that Schechtman's earlier (narrative) view offers a more suitable criterion compared to her cluster concept, and why this can be regarded as the better and preferred approach to personal identity so far. As evident by now, I am deeply troubled with the degree of importance and responsibility Schechtman ascribes to the social criterion in her cluster theory. While the flexibility Schechtman offers is commendable, it is this flexibility that leads her view into complicated territory. Having used several scenarios to demonstrate my concerns earlier, I will provide a brief synopsis here to further support my complaint. In the case where a person is in a vegetative state (having no psychological continuity) after having been involved in a horrific car accident, Schechtman's cluster theory grants society full responsibility, albeit society might not be aware, in determining whether the victim survives as the same person. It seems absurd to place such importance of identification in the hands of society who may not even be aware of their role in the victim's personal identity. It would seem equally absurd if society considers the victim as maintaining their personal identity, while Schectman's cluster concept would force us to accept that another body with the exact mental contents of the victim *also* survives as the original victim if, and only if, society regards this as the victim surviving, too. Similarly, we see how Schechtman's emphasis on social continuity in her PLV works against our intuition as witnessed in the eagle people case. In the case where I am about to be beheaded, knowledge of an herb that will alter my psychology and have me believe I will survive the beheading does not remove the deep intuitive concern I have of

whether *I* will survive the beheading, regardless of society being confident that I will survive as the eaglish entity. It is deeply counterintuitive to suggest that society has more, or even as much say about *my* survival (and that this could possibly be considered sufficient for personal identity) where I firmly believe that I will *not* survive the beheading.

Instead, Schechtman's narrative view is far more preferrable thus far, and it is this narrative view that will serve as the potential resolution, while explaining why personal identity in the sense of characterization should be what matters, not personal identity in the form of reidentification. To remind ourselves, Schechtman's (2014) narrative view explains that a person continues to be the same person over time if there exists a sense of self, or a self-understanding, where the individual has the capacity to understand thoughts, memories, experiences, and so on, as part of *their* lives. This narrative view (Schechtman, 1996) relies on the notion that a person is able to make sense of these experiences, thoughts, etc. as *their own,* and that they can place each experience, memory, and so on in the story of their lives in a way that makes sense and is meaningful, and that the person can see each thought and experience as part of *their* life. This view follows our intuition whereby it is usually the case where we accept a thought, memory, or experience is our own if we are able to fit it into our life story in a meaningful and coherent way. When confronted with an experience that we do not have any recollection of, it is not unexpected to question whether the experience really was our own experience, and not that of someone else. It would not be unreasonable to consider this narrative view as the forefront element of a personal identity theory because as an individual, we consider ourselves as having the primary and most fundamental role in the narrative of our own lives. The suggestion that Schechtman's narrative view should be primarily considered is on the basis that it works in conjunction with the concept of our consciousness, which we will be borrowing from Locke. Here I am not requiring specific memories to remain intact. Rather, for this alternative view, the narrative view requires a degree of a stream of consciousness that would in turn support a (human) being's ability to place a thought or experience into the story of their life in a logical and meaningful way. This stream of consciousness, again, will not need a person to recollect each

memory from the day before, the day before that, and so on. It will, however, serve to pave the way for a person to make meaningful sense of a thought or experience in the way that they will be in a position to recognize a specific thought or experience as their own, granted that it fits neatly into their narrative.

Furthermore, Schechtman (1996) postulates that personal identity should be tackled through the lens of characterization, not reidentification. Characterization, as described earlier, is "the question of which beliefs, values, desires, and other psychological features make someone the person she is" (Schechtman, 1996: 2). The concept of characterization works neatly with the narrative view whereby this view enables a person to attribute a thought, desire, experience, and other features to themselves, provided it meaningfully and coherently fits into the story of their life. In other words, I recognize the experience of singing in the rain as belonging to me (being *my* experience) as I can logically place it in the story of my life and the experiences I had before and after this moment. This way, we are able to ascribe features to a person with the confidence that we are ascribing these features to the *correct* person (by ensuring that the feature fits coherently and meaningfully into the person's narrative of their life), while avoiding the need reidentification theorists face of needing to explain *how* a person is the same person at t_1 one and the same person at t_2. This discussion serves to demonstrate that her narrative view, plus an emphasis on consciousness, works well as a reidentification theory.

Conclusion

Philosophers have spent significant amounts of time searching for a plausible criterion to explain why a person persists as the same person over time, without a definitive explanation being reached or agreed upon thus far. Beyond a theoretical or academic scope, the concept of personal identity carries increasing practical relevance where we are faced with scenarios, whether current or

potential, where we are required to explain who has survived, such as in real-world circumstances of amnesiac patients, or the potential of future head transplant procedures.

This discussion began with a look at what is meant by 'personal identity,' as well as what is required for a proposed criterion of personal identity to be considered. For the purpose of our discussion, personal identity as a concept was used in a metaphysical context where this construct consists of two elements, namely the self (personal or survival of oneself) and the person (being the same person to others). As the concept of survival was widely used in the above discussion, it was imperative to clarify what is meant by this notion, specifically that it is a matter of an individual entity continuing to persist over time. Finally, the four features were introduced, which include survival, moral responsibility, self-interest concern, and compensation, and were later used to examine the credibility of the six personal identity theories that were included in our assessment.

As mentioned, the contributions of several prominent philosophers were examined. Arguably one of the most recognizable theories would be that of John Locke who puts forward the notion that one's consciousness is the answer to their personal identity. In his work, Locke explains that consciousness is not synonymous with memories. Instead, the element responsible for continued personal identity over time, according to Locke, is the continuation of a person's consciousness which is supported by chains of memories. Another familiar name would be Derek Parfit who proposes a psychological criterion, namely the continuity of one's psychology. Additionally, Parfit introduces his concept of *Relation R* as the element that matters in personal survival where he describes *Relation R* as psychological continuity or connectedness where instead of possessing direct memories, a person possesses links of memories or overlapping chains of psychological connectedness.

Departing from the more psychological views, we considered an approach from Jeff McMahan whose criterion is neither distinctly psychological nor physical. Rather, McMahan's Embodied Mind View postulates that personal identity is achieved when there is present a sufficient degree of

physical continuity that has the capability to support the continuity of the capacity for consciousness. In other words, for a person to survive as the same person over time, McMahan requires that they possess their *same brain* that will serve to support the *functional continuity* that is the foundation of the *capacity* for one's consciousness (it is not the mental content which carries the importance to McMahan, rather the *capacity* for consciousness supported by the same physical brain).

Our analysis also incorporated two key physical criteria where both philosophers, namely Bernard Williams and Eric Olson, each propose their own respective version of a bodily criterion. Williams considers continued personal identity not as the presence of one's full body persisting over time, but rather as requiring *enough* of one's body to persist in a way that is able to support psychological and mental functioning. A crucial element where Williams differs from his psychological counterparts is that although his view is categorized by the need of enough of one's body to support psychological and mental functioning, Williams does not require any degree of mental or psychological content for a person to persist as the same person over time. Olson too puts forward a physical criterion suggesting that a person surviving as the same person over time is a matter of *organism* continuity. This is achieved through survived existence and functioning of the person's brainstem. The importance of the continued existence of one's brainstem, according to Olson, is that it is the structure in the body that is responsible for the organization of one's entire body that ultimately makes a person the same organism over time. A functioning brainstem is vital for enabling and maintaining basic human functioning that are imperative for sustaining life, and thus, according to Olson, is the criterion that is both necessary and sufficient for continued personal identity.

Lastly, we considered two approaches offered by Marya Schechtman, specifically her earlier (Narrative Self-Constitution) view and her later (Person Life) view. We discovered that her narrative view proposes that personal identity is explained by a person possessing a sense of self where they have the ability to construct a story (narrative) of their life in that they are able to put together the

pieces of their life in a way that is meaningful and coherent. Schechtman's later or Person Life View shifts from the narrative told from the perspective of an individual to the narrative constructed by society whereby the individual possesses a particular place in person-space and society recognizes a person to be persisting as the same person over time.

The above proposed criteria of personal identity were each considered individually and, in our analysis, we looked at the trouble each criterion faces respectively, as well as how they each fare against the four features of identity. Beginning with Locke whose approach to personal identity fails to account for significant and rational breaks in consciousness (Alzheimer's, for example). Locke's consideration of consciousness does not adequately meet the features of moral responsibility and compensation where we found it to be counterintuitive to withhold responsibility or compensation for an action if the individual in question simply has not conscious recollection of the action.

Parfit, on the other hand, finds himself caught in a web of complaints, such as scenarios where he is faced with the challenge of needing to explain who has survived in cases where it appears that a person has survived in two different places (as two individual entities) simultaneously, or where he needs to, for instance, distinguish between an individual surviving as the Martian, but that our intuition tells us that the Martian who went on to commit a crime is *not* the original Earthling who still exists in some scenarios. Parfit's psychological approach is unable to satisfy the feature of survival where there are instances (again, such as amnesia) where our intuition strongly leads us to recognize the individual in question as having survived, yet Parfit's criterion argues otherwise. Furthermore, similar to what we witnessed in our discussion of Locke, regardless of the degree of psychological connectedness (Parfit) that exists between present self and a past version of myself, there is something amiss and counterintuitive of escaping moral responsibility or compensation for an action that I committed in the past, even if it is the case that I have extremely little psychological connectedness to this version of myself. Finally, Parfit also runs into trouble with regards to the

fourth criterion where he is unable to argue for why we feel a deep sense of self-interested concern for a person that Parfit claims (and again, against our intuition) will not be *us.*

McMahan fails to provide clear argumentation in support of his (alleged) physical criterion. Rather, McMahan presents several thought experiments to demonstrate the concerns connected to other criteria of personal identity yet is unable to defend his own criterion in these scenarios. We also witnessed the charge that McMahan finds himself leaning more on a psychological criterion than he is prepared to admit. In respect to the four features, we found that McMahan's criterion of possessing enough of the same brain satisfies the feature of survival in *some* cases, but interestingly, not in all, such as what was discussed in *The Nuclear Attack.* Similarly, McMahan's criterion neatly captures the features of moral responsibility and compensation in some cases where the person still possesses enough of their own brain that maintains mental functioning. However, in a scenario such as a body swap where the persons involved no longer possess *their own* brain, McMahan is unable to explain why we intuitively believe that a *body swap* has happened and that the persons have survived in each other's bodies, despite not having their own brains. In line with this is the concern McMahan faces in connection to the fourth feature where his criterion of possessing enough of one's own functioning brain does not explain why I would be more concerned about the individual who wakes up with my psychology after a body swap procedure compared to the concern I would feel toward my original body, including what was once my brains.

One of the several concerns we explored in our discussion covering Williams's physical criterion is the trouble he faces that is similar to what we witnessed with McMahan of the scenario where the psychologies of two individuals is swapped, we intuitively consider what has happened to be a body swap procedure and that each person's respective identity follows their psychology, abandoning their original body. Williams's physical criterion of the body being the seat of personal identity not only fails to provide a suitable explanation for why we should endorse his physical criterion but is also unable to argue against our intuitive response to what we believe has happened – a body swap

(instead of a psychology swap, as Williams would like to have us believe). It was demonstrated that in some scenarios, Williams's physical criterion does not adequately meet the feature of self-interested concern, specifically in hypothetical cases where intuition tells us that our identity follows our psychology (as we saw in the thought experiment involving a psychology swap between two persons who need to decide which body receives a reward and which is tortured).

Another physical criterion we considered that is not immune to trouble is Olson's suggestion of a functioning brainstem that supports essential bodily processes. Olson faces an identical charge as Williams where as long as a person possesses enough of their functioning brainstem, a person can undergo a complete psychology change, even a psychology swap, the individual is considered to be the same person, which works against our intuition, particularly in cases where the psychology of another has been 'implanted' in the original individual's brain. We also discussed the concern of Olson's criterion facing the charge of being illogical where if a person underwent a brainstem transplant while their mental content remained intact, we would naturally be led to believe that the person survives the procedure as the same person, just with a new brainstem, which Olson would deny. It was determined that Olson's criterion of a functioning brainstem does not satisfy the feature of survival as demonstrated in the examples mentioned moments ago. It was also found to be absurd to suggest that a person is exempt from moral responsibility and compensation if they were to, for instance, undergo a brainstem transplant but maintain all the same psychological content and memories of committing a crime or earning a reward, simply because they do not possess the same brainstem. Lastly, and as seen with the previous physical criteria, Olson's criterion fails to convince us to believe that we would have more self-interested concern towards our brainstem than our psychological content.

Schechtman's later (Person-Life) view saw the introduction of her social criterion to form her cluster concept comprising of physical, psychological, and social criteria. While Schechtman grants the space of requiring only two of the three criteria, the concern with her social criterion is that it is

given too much influence. Several scenarios were discussed to expose the challenge that Schechtman's social criterion faces. In some cases where either the physical or psychological criterion is present (and not the other), the social criterion appeared to be 'a deciding factor' in determining whether a person survives as the same person. More trouble comes in when parts of society interpret who has survived in a way that differs from what others in society think, such as whether a person in a vegetative (with the absence of the psychological criterion) continues to exist as the same person – some might deem the person as having not survived, while others might have contrasting beliefs. While the social criterion has its place, there are valid concerns that it does not matter as much as the physical and psychological criteria.

Ultimately, Schectman's narrative view can be considered as providing, at least thus far, the more plausible criterion of personal identity in comparison to her Person Life View by having argued that her narrative view carries more significance than her social criterion, and that it (the narrative view) captures the four features better than her latter view. When considering the features of moral responsibility and compensation, it was discussed that the narrative view provides a more reasonable explanation for a person to coherently argue that they committed an action compared to society having as much of a say of who it was that committed an action. This argument was demonstrated using examples such as discovering the cure for cancer. I would be in a better position to have certainty that it was in fact me who discovered the cure compared to society thinking whether or not it was me. Similarly with the feature of survival. It seems clear that I would be in a better position to argue my survival than what society would be. Finally, the narrative view captures the fourth feature in a way the social element does not where if I were told about a tortuous experience I would undergo the next day, I would feel deep concern toward my well-being, despite the case where society claims it will not be *me* who will be tortured. Schechtman's narrative view, in comparison to her Person-Life View, and in conjunction with Locke's stream of consciousness, provides a more reasonable and intuitive approach to personal identity where a person is in a better

position to fits the parts of their life into their life's narrative in such a way that is meaningful and coherent.

References

Beck, S. (2014). "Transplant thought-experiments: Two costly mistakes in discounting them," *South African Journal of Philosophy,* vol. 33, no.2, pp189-199. DOI:10.1080/02580136.2014.923685.

Beck, S. (2015). "The Extreme Claim, Psychological Continuity and the Person Life View," *South African Journal of Philosophy,* vol. 34, no.3, pp314-322. Available from: https://hdl.handle.net/10520/EJC176334.

Beck, S. (2016). "Technological Fictions and Personal Identity: On Ricoeur, Schechtman and Analytic Thought Experiments," *Journal of British Society for Phenomenology,* vol. 47, no.2, pp117-132. DOI:10.1080/02580136.2006.12063060.

Beck, S. and Oyowe, G. (2018). "Who Gets a Place in Person-Space?", *Philosophical Papers,* vol. 47, no., pp183-198. DOI:10.1080/05568641.2017.1421868.

Bennet, L. (2014). *Bernard Williams on Personal Identity Thought Experiments in "The Self and the Future* [online]. Available from: https://implausibleworlds.wordpress.com/2014/07/05/williams-on-personal-identity-thought-experiments-in-the-self-and-the-future1/ [Accessed: November 2021].

Locke, J. 1975 [1694]. *An Essay Concerning Human Understanding*. Ed. P. Nidditch. Oxford: Clarendon Press.

McMahan, J. (2002). *Ethics of Killing: Problems at the Margins of life*. Oxford: University Press. https://doi.org/10.1093/0195079981.001.0001.

Melnick, S. (2017). *The Brain Behind the First 'Frankenstein' Surgery* [online]. Available from: https://medium.com/quarkmagazine/the-brain-behind-the-first-frankenstein-surgery-7eaeb1cfb5dd [Accessed November 2021].

Olson, E. (1997a). *The Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press. https://doi.org/10.1093/0195134230.001.0001.

Olson, E. (1997b). "Was I Ever a Fetus?" *Philosophy and Phenomenological Research*, vol. 57, no. 1, pp. 95–110. Available from: https://doi.org/10.2307/2953779.

Oyowe, O.A. (2010). "Surviving without a Brain: A response to McMahan on Personal Identity," *South African Journal of Philosophy*, vol. 29, no.3, pp. 274–287. Available from: https://doi.org/10.4314/sajpem.v29i3.59148.

Parfit, D. (1984). *Reasons and Persons.* Oxford: Clarendon Press. https://doi.org/10.1093/019824908X.001.0001.

Piccirillo, R. A. (2010). "The Lockean Memory Theory of Personal Identity: Definition, Objection, Response," *Inquiries Journal/Student Pulse,* vol. 2, no.8, pp. 1. Available from: http://www.inquiriesjournal.com/a?id=1683.

Schechtman, M. (1996). *The Constitution of Selves*. Ilthaca (New York): Cornell University Press. DOI: 10.7591/9781501718380.

Schechtman, M. (2010). "Personhood and the practical," *Journal of Theoretical Medical Bioethics*, vol. 31, no.4, pp. 271-283. Available from: https://doi.org/10.1007/s11017-010-9149-6.

Schechtman, M. (2014). *Staying alive: Personal identity, practical concerns, and the unity of a life*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199684878.001.0001.

Wagner, N.F. (2015). "Marya Schechtman: Staying Alive," *The Philosophical Quarterly,* vol. 65, no.258, pp.140-143. Available from: DOI: 10.1093/pq/pqu050.

Weinberg, S. (2011). "Locke on Personal Identity," *Philosophy Compass,* vol. 6, no.6, pp.398-407. DOI: 10.1111/j.1747-9991.2011.00402.x

Williams, B. (1970). "The Self and the Future," *Philosophical Review,* vol. 79, no.2, pp. 161–80. Available from: https://doi.org/10.2307/2183946.