

The Extension of Personal Identity and Personhood

by

Travis Godfrey

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

Master of Arts

In

Philosophy

Carleton University

Ottawa, Ontario

© 2019

Travis Godfrey

## Abstract

This thesis argues that personal identity and personhood can be conceived of as extended, following in line with arguments presented that claim that cognition and minds extend beyond the skin-and-skull boundary of the body. This thesis endeavors to first explicate and analyze the Hypothesis of Extended Cognition (HEC), as well as analyzing arguments that push back against this hypothesis. Arguments are provided to show that objections against HEC and the Extended-mind Corollary Thesis, the extended self thesis (EST), do not hold up. The claim is made that EST and HEC remain tenable theses. The thesis proposes to read EST through the concepts of personal identity and personhood. Argumentation is provided to support the claim that, under certain analyses, personal identity and personhood can be conceived of as extended. This is supported through the examination of several contemporary accounts of personal identity and personhood.

## Acknowledgements

Now that this thing is guaranteed to be published, I would like to acknowledge my supervisor, Professor Myrto Mylopoulos. She has provided incredible insights during my research, has been incredibly patient and supportive during this endeavor, and has succeeded in helping me understand the actual definition of the word ‘ubiquity’. Without Professor Mylopoulos’ immense kindness and gentleness, this thesis would not have been possible.

Besides my advisor, I would like to thank the rest of my thesis committee: Professor Gordon Davis, Professor Christine Koggel, and Professor Mark MacLeod, for their patience, encouragement, insightful comments, and critical analyses.

I would also like to acknowledge the following, in no particular order: Antoine Fillion, Adélaïde Caillet, Emma McBride, Katrina Rosie, Sandy Kirkpatrick, The Carleton Improv Association, and “The Adventures of Drazask, the Athlete Wizard, and his Faithful Companions”.

When I was reading a thesis acknowledgement sample, the author thanked their parents for giving birth to them. We aren’t doing that here. However, the support I have received from Selina McCarthy and Duane Godfrey throughout the past year has gone over and beyond the obligatory support a parent must give. Honorable mentions go to Ruth Booth and Justin McCarthy as well for their support and compassion.

Also, I am proud of myself for the work that I did.

## Table of Contents

<b>Chapter 1.</b>	Introduction	1
<b>Chapter 2.</b>	The Hypothesis of Extended Cognition	5
1.	Introduction	5
2.	The Hypothesis of Extended Cognition	7
2.1.	Adams and Aizawa	13
2.2.	Rupert	19
2.3.	Sprevak and Miyazono	25
2.4.	Milojevic	34
3.	Conclusion	40
<b>Chapter 3.</b>	The Extended Self Thesis and its Objections	44
1.	Introduction	44
2.	Preliminaries	46
3.	The Extended Self	53
4.1.	Contextualization: Olson	56
4.2.	Contextualization: Baker	59
4.3.	Contextualization: Wilson and Lenart and Heersmink, Oh My!	60
5.	Defenses against EST	63
5.1.	Olson	63
5.2.	Baker	72
6.	Conclusion	79
<b>Chapter 4.</b>	Extended Memory	82
1.	Introduction	82
2.	Preliminaries	86
3.	Memory and Addis and Tippet	90
4.1.	Extended Memory	102
4.2.	Non-Autobiographical Extended Memories	104
4.2.	Autobiographical Extended Memories	108

5.	Conclusion	116
<b>Chapter 5.</b>	<b>Extension of Personal Identity and Personhood</b>	<b>118</b>
1.	Introduction	118
2.1.	The Person Life View	119
2.2.	Milojevic	134
2.3.	Olson and Williams	136
3.	Parfit	141
4.	Conclusion	148
<b>Chapter 6.</b>	<b>Conclusion</b>	<b>151</b>
<b>Works Cited</b>		<b>157</b>

## Chapter 1. Introduction

In the following chapters, this thesis will endeavor to investigate an implication inferred in the 1998 paper by Clark and Chalmers. The authors propose a functionalist reading of the mind, specifically that we can conceive of the mind as extending beyond the body, the skin-and-skull boundary. This hypothesis has been widely discussed, remaining a contentious postulation on the nature the mind and of cognition. The immediate implications of this thesis are daunting. If we allow that cognition and mental events extend beyond the body, then we can conceive of the externals as altering and expanding the mind. This claim does not have to be as radical as initially perceived. In general, tools extend cognition. Use of a tool, even temporarily, changes the neural mappings of the brain. For example, the receptive visual field of a macaque using a rake for as little as thirty seconds becomes elongated as if the rake were part of the arm (Clark 2005), or like how a walking stick extends sensorimotor perceptive capabilities of a blind person (Vold 2015). Or in another way, if we allow that the mind extends beyond the body, then it seems that non-organic externals can aid in the constitution of mental states. Furthering this idea allows the possibility of there being non-human or non-biological entities that can constitute mental events. Advancing this thesis logically can bring about even more intimidating implications. If the thesis is true, then there seems to be nothing immediately sacred about the brain when identifying minds. Sure, the brain at this moment is the only thing that has the causal architecture to instantiate mental states. However, if we follow Clark and Chalmers in identifying cognition and mental states by their functional description, then it seems that we can imagine entities who instantiate much of their mental life by virtue of technological artifacts. It may not be that an individual brain alone thinks, and instead that it is the hybrid of human brain and external that is responsible for the mental states of the human. The thesis will provide an

exposition of why Clark and Chalmers believe that we can conceive of the mind as extended in the next chapter.

While the extended mind is a topic of this thesis, it will remain but a preliminary focus to a larger topic. Clark and Chalmers, in addressing the feasibility of the extended mind thesis, allude to the possibility of extended agents, or extended selves. They believe that the proper way to conceive of agents is as spread out into the world, rather than as restricted and limited by the skin-and-skull boundary. It is this particular thesis, the extended self thesis (EST), that I wish to explore. Specifically, I want to bring in a topic that has fascinated and inspired me in philosophy: I want to read EST through the lens of personal identity and personhood. I believe that it is an apt endeavor to attempt to understand the EST through the concepts of personal identity and personhood. EST is concerned with the question of whether we extend beyond our bodies, and the literature of personal identity and personhood tries to establish exactly what we are: what are our persistence conditions, what makes a person (a self) what they are and what ensures that they continue to exist, how do we individuate persons from other persons or identify a non-person from a person, etc. I argue in later chapters why reading EST through personal identity and personhood is a pertinent and advisable endeavor.

This project will proceed as follows. This thesis will begin by expositing Clark and Chalmers initial arguments for why cognition extends, and subsequently why we can think that the mind extends. Following this will be an analysis of a few prominent authors who argue against the hypothesis of extended cognition (HEC). I focus on authors that disagree with HEC for the following reason: HEC has been thought of as grounding later arguments for the extended mind, and then later for the extended self. If one can disprove the possibility of extended cognition, then the reasons for thinking that the mind extends or that the self extends are less tenable. Thus, I focus on analyzing and responding to criticisms levied against HEC.

Following this, the thesis attempts to expost the original claim given by Clark and Chalmers in favour of viewing the self as extended. Following the conclusion that this thesis will endeavor to analyze EST through personal identity and personhood, we will move on to analyze a few authors who reject the possibility of EST. Argumentation will be given to claim that these authors reject EST hastily, and that there is still room to argue for the tenability of this thesis. I will provide an example of one way of viewing personal identity and personhood as extended, namely through the criteria of memory. Argumentation will be given to show why memory can be seen as an important criterion for various analyses of personal identity and personhood, as well as arguments by other authors that advance the claim that memory can extend beyond the body. Following this, a few analyzes of some prominent accounts of personal identity and personhood will be assessed to gauge whether they can be accommodated with extension. Ultimately, the conclusion of this investigation will be that extension's tenability is contingent upon the particular analyses of personal identity or personhood that we are operating with. Thus, while this thesis has provided good reasons to think that certain analyses of personal identity and personhood are accommodatable with extension, this thesis does not endeavor to definitively prove that personal identity and persons extend. This would require the thesis to posit a view of personal identity and personhood as the proper analyses of said concepts. Providing true analyses of personal identity and personhood is beyond the scope of this thesis, and as such the conclusion will be a bit less assertive in order to maintain a conservative stance on what we can adequately and justifiably argue for in this thesis. This thesis will conclude with a discussion of a problem facing the extension of personal identity and personhood, as pulled from Derek Parfit. Finally, this study will conclude with a limited discussion of how this thesis might impact upon theories of collective entities and macrocognition.

The next chapter will begin our investigation by expositing Clark and Chalmers' 1998 article, "The Extended Mind", as well as discussing a few authors who disagree about the boundaries of cognition.

Also, if your name is Shelton, I greatly apologize and did not mean to provide an unnerving caricature of you.

## Chapter 2. The Hypothesis of Extended Cognition

### 1. Introduction

Andy Clark and David Chalmers advance a thesis of active externalism in their joint article “The Extended Mind” (1998). Externalism and internalism in the philosophy of mind are views about the processes and vehicles of cognition but disagree about the boundaries of these vehicles. Internalism about the mind claims that the boundaries of the mind are within the physical body of the organism, the ‘skin and skull’ as Clark and Chalmers put it. Externalism allows that the boundaries may extend beyond the organism itself. Likewise, externalism can allow for cognition to be embedded in the environment, as cognitive processes are recognized to depend very heavily, “on organismically external props and devices and on the structure of the external environment in which cognition takes place” (Rupert 393). Active externalism, as advanced by Clark and Chalmers, is the claim that the environment can play an active role in driving cognitive processes, i.e. active in the sense that cognitive processes aren’t limited to the brain and body on this account but can extend into the environment. “The relevant external features are active, ‘in the loop’, insofar as the external features are just as causally relevant as typical internal features of the brain” (Clark and Chalmers 9). This thesis of active externalism about cognitive states will be referred to as the Hypothesis of Extended Cognition (HEC).

Differing authors have articulated this thesis in different ways: Rupert (2004), for example, understands HEC as follows: “human cognitive processing literally extends into the environment surrounding the organism, and human cognitive states literally comprise—as wholes do their proper parts—elements in that environment” (389). Adams and Aizawa (2001) understand what they call the ‘transcorporeal’ or ‘extracranial’ position as the view that if “cognition is simply computation over representational states, and if one’s tools such as paper

and pencil, form or contain representations, then one has a case for the radical view that, in at least some cases of tool use, cognition extends beyond the boundary of the brain” (46). Clark (2011) himself can lend to the understanding of HEC in claiming that, “under certain conditions, [nonbiological] props and structures might count as *proper parts of extended cognitive processes*” (2011, 68 author’s emphasis). The initial claim for extension, albeit not articulated as the Hypothesis of Extended Cognition, is the claim that “if, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process” (Clark and Chalmers 1998, 8). For the most part, each of the cited authors define HEC in relatively similar ways. However, since there are areas of dissimilarity in their views<sup>1</sup>, this thesis will go forward with the understanding of HEC as claiming that cognitive processes can and do extend beyond the body. The reason for defining HEC in this way is to provide a broad definition that will be understood through an exposition of Clark and Chalmers’ initial arguments for this thesis.

This chapter will endeavor to exposit Clark and Chalmers’ initial argument for HEC and the extended mind, as well as to introduce the implications their arguments have for our conception of sociality and selves. Following this will be an investigation of some prominent authors who refute the possibility of HEC. Through a detailed exposition and analysis of their views, it will be concluded that these arguments are not enough to definitively rebut the possibility of HEC. The chapter concludes by addressing the arguments of an author that both salvages HEC and proposes a modest Extended Self Thesis (EST), leading us towards the later

---

<sup>1</sup> For instance, Rupert addresses human cognition while Clark defends in his later works a notion of cognition uninhibited by “anthropomorphic” limitations (See Clark 2005, 2011), and Adams and Aizawa’s definition already entails their conception of cognition as entailing a computational structure.

goal of this thesis of analyzing and arguing for the tenability of EST. What follows is an exposition of Clark and Chalmers' arguments for HEC as proposed in their 1998 article.

## 2. The Hypothesis of Extended Cognition

In their larger argument for the extended mind thesis, Clark and Chalmers intend to show that external traces (like pencil marks in a notebook) can be considered among the physical vehicles of specific dispositional beliefs. It is not that external, passive coding might somehow behave exactly like the fluid, automatically responsive resources of biological internal memory. Rather, it is that, "external encodings were, under certain circumstances, capable of becoming so deeply integrated into online strategies of reasoning and recall as to be only artificially distinguished from proper parts of the cognitive engine itself" (Clark 77). Clark and Chalmers make use of the game Tetris in order to help motivate their thesis of active externalism. They provide three examples: the first is of a person in front of a computer screen which displays images of various two-dimensional geometric shapes and is asked to answer questions concerning the potential fit of such shapes into depicted 'sockets'. To assess the fit, the person must mentally rotate the shapes to align them with the sockets (Clark and Chalmers 1998). The second is a person who now has an apparatus that can physically rotate the image (let's call this technology a 'button') on the screen. The third is a person, placed in a "cyberpunk future", who has a neural implant which can perform the rotation operation. Clark and Chalmers suggest that all three cases involve similar computations. The first example describes a case of mental rotation, while the second looks like a case of nonmental rotation. The third example seems to be on par with the first example, while the second example displays the same sort of computational structure as the third. If the third example is deemed to be cognitive, then on what grounds, argue Clark and Chalmers, do we claim that the second is fundamentally different? A functionalist argument is motivating the analogies in this first example, the claim

that if a state plays the same causal role in the cognitive network as a mental state, then there is a presumption it is the mental state (Chalmers Forthcoming). In examples like an individual using a 'button' to rotate images or the re-arrangement of letter tiles to prompt word recall in scrabble, the individual brain performs some operations while others are delegated to manipulations of external media.

Building on the idea that extended cognition is a core cognitive process, Clark and Chalmers want to advance a further thesis, namely that the mind itself is generally an extended entity, an entity with parts outside the head. The extended mind thesis can be roughly summarized as arguing that the technology and environment we use can become a part of our minds, that is, can be partially constitutive of our mental states. "Proponents of the extended mind story hold that even quite familiar human mental states (e.g., states of believing that so and so) can be realized, in part, by structures and processes located outside the human head" (Clark 2011 76). The central claim made is that if they can show that beliefs can be constituted by features of the environment, in that those features play the right sort of role in driving cognitive processes, then other mental states (and minds as a whole) extend into the world.

Clark and Chalmers argue for this through the case of Otto and his notebook. First, they illustrate a paradigm example of belief embedded in memory. A woman named Inga hears about an exhibition at a museum and decides to go see it. She recalls that the museum is on 53rd street and walks there. It seems clear that Inga believes the exhibition is on 53rd St., and that she believed this before going to the museum. The belief of the location of the museum was not an occurrent belief but was somewhere in memory, waiting to be accessed. They then introduce Otto, an individual suffering from something akin to Alzheimer's disease. Because of a failing faculty of memory, Otto relies on information in the environment to help structure his life. Otto has a notebook that he uses to record and access information, thus his notebook plays

the role usually played by a biological memory. He hears about the exhibition and decides to go. He accesses his notebook which states in its contents that the museum is on 53rd St. Clark and Chalmers posit that these two cases are entirely analogous, as the notebook plays the same role that memory plays in this example: “The information in the notebook functions just like the information constituting an ordinary non-occurrent belief; it just happens that this information lies beyond the skin. Insofar as beliefs and desires are characterized by their explanatory roles, Otto’s and Inga’s cases seem to be on par” (13).

Before moving on, it is important to note a few important definitions. The ‘Parity Principle’, as articulated by Clark and Chalmers, is that if, “as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process” (8). The Parity Principle is their main motivation for thinking that Otto’s beliefs extend, as well as for the larger general reasoning that the mind extends. The Parity Principle allows, as Clark and Chalmers argue, for treatment of the examples that deal with mental rotation of Tetris inspired images as cognitively on par with one another.

Also, it should be noted that Clark and Chalmers list four general grounds for ascribing an extended belief (specifically in reference to Otto): 1- The notebook is a constant in Otto’s life—in cases where the information in the notebook would be relevant, he will rarely take action without consulting it. 2- The information in the notebook is directly available without difficulty. 3- Upon retrieving the information from the notebook he automatically endorses it. 4- The information in the notebook must be consciously endorsed at some point in the past, and indeed is there as a consequence of this endorsement (Clark and Chalmers 15-16).

Clark, in a separate article (2005), proposes the Glue and Trust conditions: 1- That the resource must be reliably available and typically invoked. 2- That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny. Additionally, it should be deemed to be as trustworthy as something retrieved from biological memory. 3- That information contained in the resource should be easily accessible as and when required (Clark 2011, 46). All of the above are presented to aid us when deciding whether the claim to cognitive equivalence is justified or not. Thus, when reflecting upon HEC or upon the extended mind thesis, a functionalist comparison is part of the realization of extended states. These conditions and principles are thus constraints on the functionalist argument: these are all proposed to allow us a framework for recognizing when nonbiological candidates are to be seen as included in an individual's cognitive system, and thus allow a modest HEC (in contrast to a radical HEC that will be discussed in 2.3).

Upon proposing the conceptual possibility of the extension of the mind, Clark and Chalmers briefly discuss the implications this has for thinking about cognitive agents. With the acceptance of the Extended-Mind Thesis comes the possibility of socially extended cognition; mental states constituted in part by the states of other thinkers. Others may try to make sense of the possibility of socially extended cognition by invoking transactive memory. Certain social groups can develop types of distributed memory systems where there is a cognitive interdependence between group members, which means that agents rely on each other for their cognitive performance (Heersmink 3137). During transactive encoding, group members discuss incoming information and negotiate who will store the information and in which form. During transactive retrieval, group members have to determine where the information is stored, so as to integrate data stored in different brains (Haris et al. 2010). Heersmink provides an example to show that transactive retrieval and encoding need not require that one individual is

the designated processor, but rather that transactive memory can be jointly stored and recalled. In the example, two individuals are attempting to remember a show that they watched. Through a series of prompts and negotiation they are able to deduce the name of the show (see Heersmink 2017).

Likewise, Wegner et al. (1985) emphasise the interdependence of individual memory systems and therefore point out that transactive memory systems cannot be reduced to individual memory. In this way we can think of these transactive memory systems as a group-level property that exists only when people interact and communicate in the right way (Heersmink 3137).

The notion of socially extended cognition is widely supported<sup>2</sup>. Baker notes how spouses can serve as vehicles for memories, and, more generally, how individuals can act to support memory. Wilson and Lenart advance that memory can be constructed by external individuals passively through reification, or constitutively through active participation. Gallotti and Huebner argue for a stronger thesis of socially extended cognition, showing that our understanding of the world is determined partly by 'what's in the head' and partly by linguistic and social conventions that are shared by the members; much of cognition is socially situated and socially scaffolded (252). Cash (2010) utilizes similar implications of collective or extended cognition and agency to make sense of collective and individual responsibility. He claims that HEC critiques individualism in cognition and uses these critiques in line with feminist critiques of individualist accounts of self, agency and autonomy in order to reconceptualize cognition and concerns about responsibility as fundamentally communal, social, and normative (645). He argues that, with the

---

<sup>2</sup> See Heersmink (2016; 2017), Baker (2009; 2013), Cash (2010), Wilson (2004), Wilson and Lenart (2014), Gallotti and Huebner (2017)

acceptance of HEC, one can see individuals as merely one of many possible loci of cognition, action, and responsibility.

Clark and Chalmers also advance a seemingly more controversial possibility, which will be central to this thesis; the possibility of extended selves. “Does the extended mind imply an extended self? It seems so. Most of us already accept that the self outstrips the boundaries of consciousness; my dispositional beliefs, for example, constitute in some deep sense part of who I am. If so, then these boundaries may also fall beyond the skin” (Clark and Chalmers 18). The information in Otto’s notebook seems to be a central part of his identity as a cognitive agent. The authors argue that the best way to regard Otto himself is as an extended system, a coupling of biological organism and external resources. They argue that to resist this conclusion would entail shrinking the self into a mere bundle of occurrent states, severely threatening its deep psychological continuity (18).

As has already been noted at the outset of this project, a primary focus will be the Extended Self Thesis (EST). However, in conceptualizing this thesis it is necessary to understand the framework that helped conceive of it. As such, it is necessary to understand the foundations of HEC in discussing the radical implications of EST. However, before analyzing what EST entails, it will be advisable to at least consider HEC in some detail. The thesis is not without objections. In what follows I will briefly lay out what I consider to be some of the strongest objections to HEC. However, it will not be the goal of this thesis to defend HEC from these objections, but rather to articulate HEC and the ways it has evolved (i.e., has been responded to) in the face of these objections. To provide a systematic defense of HEC against all of the objections that are levied against it is beyond the scope of this thesis<sup>3</sup>. My investigations of EST will take HEC for

---

<sup>3</sup> See Sterelny 2004, Gertler 2007, Drayson 2010, Walter 2010, Wheeler 2010, etc.

granted. Insofar as EST presupposes HEC, an investigation will be undertaken to analyze the feasibility of the latter thesis. For, if it turns out that cognition cannot conceivably extend, then it seems that considering a self as a diffuse and extended entity may be to mischaracterize the notion of selves<sup>4</sup>.

## 2.1. Adams and Aizawa

Adams and Aizawa (2001) respond to the 1998 paper, taking the tact of claiming that Clark and Chalmers mischaracterize tool-use and external artifact utilization as cognition: what they call extended cognition can be surmised as simply intracranial cognition aided by tool use. They argue that, “as a matter of contingent empirical fact, in all actual cases of human tool use brain-bound cognitive processes interact with non-cognitive processes in the extracranial world” (46). While they do not deny that transcranial cognition (extended cognition) is possible, they think that, “enough is known about psychological processes and other physical processes to rule out the possibility that in, in real world cases, we have cognitive processes spanning the bounds of the brain” (46). The authors note that ‘being inside the brain or skull’ cannot be the mark of the cognitive. As such, they endeavor to identify ‘the mark of the cognitive’ and note the types of processes in the world that have that mark. Through this, the authors claim the conclusion that only intracranial processes count as cognitive. It appears to be just a contingent empirical fact, they claim, that cognitive processes are not transcorporeal processes (48).

Adams and Aizawa posit two important conditions of the mark of the cognitive. The first is that cognitive states must involve intrinsic, non-derived content. “So, it seems to us, intrinsic

---

<sup>4</sup> I use ‘self’ here loosely. A self can entail personal identity, personhood, soul, etc. As will become relevant in chapter three onwards, I aim to discuss EST under very specific parameters. For the time being, EST will not be investigated in great length and as such ‘the extended self’ here refers to the thesis referred to by Clark and Chalmers and other authors who critique HEC.

content is a legitimate necessary condition on a state or process being cognitive” (50). The authors state that strings of symbols or numerals of various sorts represent what they do in virtue of conventional associations between them and social agreements and practices. Cognitive states in normal cognitive agents, by contrast, “do not derive their meanings from conventions or social practices...it is not by anyone’s convention that a state in a human brain is part of a person’s thought that the cat is on the mat” (51). Non-derived content, or intrinsic content is to be contrasted with, “for example, the way a public language symbol gets its content by conventional association” (Clark 90). Even though non-derived content is necessary for a state or process to be cognitive, Adams and Aizawa admit that it is unclear to what extent each cognitive state of each cognitive process must involve non-derived content (50). However, given that cognitive states can be to some extent less than maximally dependent on non-derived content, i.e that cognitive states can derive from both derived and non-derived content, it seems possible that this concession can open up the possibility of, “inserting extracranial states and processes into cognitive processes” (Adams and Aizawa 51).

The second condition of the mark of the cognitive is that cognitive processes are causally individuated. Adams and Aizawa claim that the cognitive may, “be assumed to be like other natural domains, namely, the cognitive must be discriminated on the basis of underlying causal processes” (52). This follows from the claim that some mechanisms will count as cognitive while other mechanisms that might suffice will not (52). The gist of their argument comes from the presumption that the purpose of science is to ‘carve nature at its joints’; science tries to get beneath observable phenomena to find the real causal processes underlying them; and tries to partition the world into causally homogeneous states and processes (54). What this means for cognition is that, if we are to be true to scientific goals of partitioning nature, then we ought to see cognition as a natural kind that can be individuated. By Adams and Aizawa’s

account, if cognition is a natural kind then there is a characteristic set of causal processes, “found by painstaking empirical investigation, to pervade the internal, biologically supported aspects of human cognitive architecture” (Clark 92).

The authors believe that their description of the cognitive allows for them to avoid the functionalist conclusion of Clark and Chalmers. Their idea is that, when one looks at the processes which embed non-derived representations, such processes happen to occur almost exclusively within the brain. Recalling the three Tetris examples, Adams and Aizawa claim a disanalogy between the second example (the man using the ‘button’) and the first in that, “the process that physically rotates the image on the screen at the push of the button as described in case (2) is not the same as the cognitive process that occurs in the brain” (54). This sort of causal process is argued to not be the same as any cognitive process, or any fragment of a causal process, in the brain. Likewise, Adams and Aizawa suggest that, with the case of Otto and Inga, that there are important and relevant differences. One is that the symbols written in Otto’s notebook have merely derived content, whereas the recollection in Inga’s brain has non-derived content. Additionally, the authors argue that Inga and Otto carry out distinct processes in coming to arrive at the museum. Otto’s “memory recall” involves picking up the notebook and turning to the appropriate page in the notebook. This involves processes that have no analogue in Inga’s memory recall. Adams and Aizawa claim that it is unreasonable to say that Otto’s “memory recall” involves cognitive-motor processing found in Inga’s memory recall. In addition, Otto’s “memory recall” involves visual processing for turning to the appropriate page of the notebook and reading the address. Inga’s memory recall does not. Further, Inga’s memory recall uses some capacity of the brain that Otto has lost due to his Alzheimer’s disease. It is because of these disanalogies that we cannot aptly conceive of Otto’s notebook as constituting part of a

cognitive system: the notebook does not make use of non-derived content and the underlying causal processes are different enough so as to not entail a shared natural kind.

This is the larger form of how the rest of their arguments against HEC are presented. By presenting various aspects of human cognition in conjunction with supposed examples of HEC the authors show how their conditions on the cognitive entail certain disanalogies. Granting Adams and Aizawa's claims about the cognitive, it is still possible to push back against their claims against HEC. First, let us analyze the first condition of the mark of the cognitive, that cognition involves particular kinds of processes involving non-derived representations. Why is it that cognition must involve non-derived content? Is it not possible that allowing the concession that cognitive states can be, to some extent, less than maximally dependent on non-derived content leads to the empirical possibility of having cognitive states constituted by derived-content?<sup>5</sup>

However, for the sake of argument, this condition of the mark of the cognitive will be accepted at face value. Even with the acceptance of cognition as involving non-derived content, it is possible to save HEC. Adams and Aizawa claim that, "whatever is responsible for non-derived representations seems to find a place only in brains" (63). Otto's inscriptions in his notebook do not seem to involve non-derived content. The notebook contains passive representations that are derived from public and social convention. However, one need not give up on HEC here unless one accepts that a proper cognitive system cannot trade in conventional representations in any of its parts. Remember that Adams and Aizawa posit that there is the empirical possibility that a cognitive system can involve derived content. When thinking about

---

<sup>5</sup> And with this, why not think that representationalism (cognition involves representations having non-derived content) is not equally possible with this caveat, a thesis Adams and Aizawa rejected in their analysis.

Otto and his notebook, the cognitive processes that underlie his occurrent belief do trade in derived representations, but they also are constituted by non-derived content within Otto.

Clark (2005) offers a thought experiment that helps elucidate this point. Imagine some Martians endowed with an extra biological routine that allows storage of bitmapped images and important chunks of visually encountered text. It would not be a large step in embracing that this storage is part of the Martian's cognitive equipment, even though it is a conventional form of external representation. Clark argues that if we accept the Martian memory as cognitive, then it follows that we ought to extend the same courtesy to Otto. "Thus, even if we demand the involvement, in any cognitive process, of at least some items that bear their contents intrinsically, it is quite unclear how we should distribute this requirement across time and space. From the requirement, if it is a requirement, that every truly cognitive agent trade in states that bear intrinsic contents, it cannot follow that every proper part of the cognitive system of an agent must trade (and trade solely) in such contents" (Clark 92).

In regard to the second mark of the cognitive, remember that part of the motivation for recognizing cognitive processes as being discriminated on the basis of underlying causal processes is to help establish cognitive processes as a natural kind. The recognition of cognitive processes as being determined on the basis of causal processes concerns, "the possible existence of a characteristic set of causal processes found to pervade the internal, biologically supported aspects of human cognitive architecture" (Clark 2011 92). The kinds of laws and regularities the authors address inform us about some commonalities in human biological memory (like priming effects) and illustrate features that reflect the detailed operation of processes internal to the brain. Clark (2011) writes that, "[s]ince these clearly pertain to some of our paradigm cases of terrestrial cognition, we should believe that these kinds of causal process are essential to the cognitive" (93).

Rupert (2004) raises a similar worry, arguing that extended 'memory' states (processes) differ so greatly from internal memories (the processes of remembering) that they should be treated as distinct kinds, quelling any temptation to argue for HEC from brute analogy. The example he gives is of a subject who is given a list of name pairings and told to memorize them. They are then tested on their rote memorization. The next step involves swapping the associations around, so that an A-B association now is an A-C association. Again the subjects are asked to memorize the list and then tested. The result is a *negative transference*, where the subjects consistently remembered less on the second step than on the first (as induced by their attempt not only at memorizing new associations but also of working against existing associative structures). With the utilization of extended memory (writing down the associations), Rupert suggests that, "there is no reason to expect negative transfer in the learning of paired associates when a subject relies on an external store" (414). Granting this significant difference, Rupert thinks it advisable to think of the two memory systems as two different kinds.

There are reasons to think that this criticism does not hold up. Rupert is claiming that there is no reason to expect negative transfer in the learning of paired associates when a subject relies on an external store. But it is unclear why this is a problem, since the cognitive processes involved in an instance of extended cognition may not necessarily induce all of the cognitive processes involved in bare internal cognition; two sets of cognitive processes need not be identical. In an attempt to memorize the pairings, we can explain the negative transfer through associative thinking. The subjects formed associative structures to memorize the pairings. Thus, when the pairings were changed it was not only a matter of memorizing the new pairings but also counterconditioning the pre-existing associative structures. When thinking about extended memory in this instance, it is not clear that cognitive associative structures are formed. The subject is recording listings by a confederate. Since they can offload some of their computations,

the subject need not incorporate all of the internal cognitive processes necessary in their current task. Clark responds in a similar way, through allusion to his Martian functionalism: “To insist that some alien mode of storage and retrieval was not cognitive just because it failed to exhibit features such as recency, priming, and crosstalk would be simultaneously to scale new heights of anthropocentrism and neurocentricism, inflating properties of the human neural realizers of certain brainbound cognitive process into requirements that must be met before any process is properly deemed cognitive” (93).

Adams and Aizawa worry about the nature and feasibility of the scientific enterprise implied by taking ‘transcranialism’ seriously. In that, by trying to find natural kinds by carving nature at its joints, transcranialist cognition will prove to have little or nothing in common by way of underlying causal processes. However, it is not clear that acceptable forms of unification require that all the systemic elements behave according to the same laws. Likewise, it seems possible that the inner goings-on that Adams and Aizawa claim to be paradigmatically cognitive themselves will turn out to have little in common as detailed causal mechanisms go. Clark contends that the argument for scientific kinds is doubly flawed: “It is flawed by virtue of its rather limited conception of what makes for a proper scientific or explanatory enterprise. And it is flawed in its assessment of the potential for some form of higher-level unification despite mechanistic dissimilarities. It is, above all else, a matter of empirical discovery, not armchair speculation, whether there can be a fully-fledged science of the extended mind” (95). Thus, the argument from scientific kinds should not be seen as a hinderance.

## 2.2. Rupert

Though I have already responded to one objection that was raised by Rupert (2004), there are still further challenges that he raises for HEC. Ultimately Rupert argues that HEC

implies highly counterintuitive attributions of belief and that, as a causal-explanatory hypothesis, HEC appears to be of marginal interest as part of a philosophical foundation for cognitive science.

Rupert motivates his rejection of HEC through the introduction of the Hypothesis of Embodied Cognition (HEMC), a thesis that claims that cognitive processes depend very deeply, “on organismically external properties and devices and on the structure of the external environment in which cognition takes place” (393). Rupert sees HEMC and HEC as offering distinct and competing explanations of various cognitive phenomena. Ultimately, Rupert aims to show that HEMC offers the superior explanation over HEC in most cases, thus prompting the endorsement of HEMC over HEC, as well as to demonstrate that HEC does not provide a promising framework for the pursuit of cognitive science. Effectively, Rupert argues along a similar vein as Adams and Aizawa: that once one considers the fine-grained functional structure of certain cases, one can see that actual extended processes are not functionally like any internal cognitive process.

One criticism of Rupert’s is that the four general conditions Clark and Chalmers posit in the general ascription of extension to a mental state actually lend themselves to HEMC rather than HEC. It bears remembering that the fourth condition, that information in the notebook has been consciously endorsed at some point in the past, and indeed is there as a consequence of this endorsement (Clark and Chalmers 17). Although this condition saves HEC theorists from absurd ascriptions of belief (Rupert 404), a consideration speaks against the HEC theorist’s acceptance of this past-endorsement criterion. The past-endorsement criterion undermines one of the important theoretical implications of HEC, “that there is no good reason to assign special status to the boundary between organism and environment. If an extended (or any) belief requires conscious endorsement in order to be a genuinely held belief, and conscious

endorsement is ultimately an internal process (that is, one that takes place within the organismic boundary), then the traditional subject is privileged in a deep sense, after all" (404). The acceptance of the past-endorsement criterion comes not only with the worry mentioned above, but also with a further reason to prize HEMC over HEC. There is less of a reason to view externals as anything more than tools used by the mind, as opposed to parts of it (this follows from allowing that only internalist endorsements are what matter in deciding an external as a part of the cognitive). "Given that internal consciousness provides the ultimate source of cognitive authority, it seems quite natural to say that the thinking subject, traditionally conceived of, is *using* those external resources" (405).

Granting that a past endorsement was required, Rupert does not identify exactly what it is that made the endorsement, instead just stating that it must have been some mechanism that can only be realized internally, as it is only internal structures that can realize conscious endorsement. This seems like a basic thing to grant; it might be the case that conscious endorsement can only be accommodated by an internalist position (as of now) insofar as we cannot attribute the processes necessary for conscious endorsement to an external. This does not mean that the proper parts of a cognitive system itself are not part of the processes necessary to entail the endorsement. Clark, in talking about the appeal to the coupling of organism and object, says that, "the appeal to coupling is not intended to make any external object cognitive. Rather, it is intended to make some object, which in and of itself is not usefully (perhaps not even intelligibly) thought of as *either cognitive or noncognitive*, into a *proper part of some cognitive routine*" (87, author's emphasis). Thus, the response is to admit the internal privileging as a contingent fact. First off, yes. It is the internal structures that are typically and usually responsible for conscious endorsement. However, when we identify an external as being part of an organism's cognitive structures, these external processes can aid in the actualization

of conscious endorsement itself. In this way, while much of the processes needed for conscious endorsement are internal, they can be supplemented by processes external to the organism.

This response might do more harm than good. One could respond by saying that this notion would create situations of confusion in regard to where conscious endorsement actually lies. At early enough points in an organism's life it would necessarily be without externals with which it is coupled or finds its processes extended with. How would the organism's initial conscious endorsement be any different than subsequent endorsements? Or say we talk of a 'veteran' extended entity, who has consciously endorsed many externals and also has processes of conscious endorsement aided by externals. Say the individual begins a process of conscious endorsement but is not aided by all externals that normally aid in some processes of conscious endorsement. Is this conscious endorsement less viable in that it is not making use of all processes that would aid in conscious endorsement? What happens if it is possible that two externals, when made available, would work towards opposite endorsements?

The second possibility is to claim that internal privileging that seems to follow from the condition of past-endorsement is only a contingent fact as of now. It is because there currently do not exist any externals that can aid or actualize conscious endorsement fully that conscious endorsement remains an internalist operation. This, like the first maneuver, is speculative. Regardless of this, a possible defense can be given by the HEC theorist that may be enough to dissuade one from thinking that HEMC is better equipped to explain cognitive extension in this case than HEC. In responding, there is always also the possibility of revising this condition or getting rid of it. However, getting rid of this condition can lead to radical ascriptions of HEC, as will be noted in section 2.3.

Rupert notes that HEC theorists insist that they can offer HEC as an explanatory hypothesis in cognitive science. One central claim is that HEC can provide taxonomies that include overarching cognitive kinds that provide the most empirically powerful framework for research in cognitive science. Rupert challenges this assertion by focusing on one specific kind of cognitive state that HEC theorists claim as possibly extended and internal. Rupert argues that, “the external portions of extended “memory” states (processes) differ so greatly from internal memories (the process of remembering) that they should be treated as distinct kinds” (407). Rupert believes that this analysis will delegitimize arguments for HEC from brute analogy. In this way Rupert’s criticism mirrors Adams and Aizawa’s. The three authors are effectively arguing that, on a fine-grained level, extended processes are functionally unlike internal cognitive processes. Thus, they should not be deemed cognitive as they are substantially of different kinds. One of the ways Rupert argues for this claim has already been addressed in the previous section; through the lack of negative transfer in external memory. Other arguments he gives in favor of disanalogy reference malfunction<sup>6</sup> and the diversity among internal memory systems, a claim, even if there is some diversity among internal memory systems, that, “there is also evidence of substantial coherence among these systems—a coherence that renders them closer in kind to each other than anyone is to what are typically claimed by HEC theorists to be external memory systems” (419).

Finally, Rupert makes the claim that most HEC motivated arguments contain a clear functionalist strain, and that, “the analysis of common-sense concepts of cognitive states does not support HEC, for common sense rules strongly against external portions of memories and

---

<sup>6</sup> He writes that “[s]ome proponents of HEC have recognized that, when faced with a malfunctioning extended cognitive system, it is useful to distinguish between the organism and the environment as separate components of that system. We seem forced, then, to recognize two different explanatory kinds, internal memory and external resources used as memory aids, with no reason yet found to think that external aids constitute genuinely cognitive states or processes” (416).

other cognitive states” (422). The functionalist approach is to yield functional-role descriptions of mental or cognitive states from common-sense psychological concepts. However, Rupert argues that this approach fails as the analysis of these common-sense concepts will not support actual common sense itself in thinking about extended processes and mental states.

The first of these two challenges by Rupert can be easily responded to. One option is to argue that Rupert himself has gotten common-sense intuitions about cognitive states and mental processes wrong. Or, that while contemporary common-sense does not support the analysis of these concepts through extension, future attitudes will be able to as technology evolves and we come to better understand the cognitive. However, a stronger move is through reference to Sprevak (2009) and Clark’s allusion to the Martian analogy. The underlying intuition is that it is possible for creatures with mental states to exist even if such creatures have a different physical and biological makeup than ourselves. There is no reason why an intelligent Martian should mirror our fine-grained psychology exactly. For instance, they might have a different physical and biological makeup than ourselves. Or specifically, Martians could possess a biological routine that allowed them to store bitmapped images of important visually encountered text. If we allow this possibility, then Clark claims that, “we would have no hesitation in embracing that kind of bitmapped storage as part and parcel of the Martian cognitive equipment. But what is stored is just a bitmapped image of a fully conventional form of external representation. If, courtesy to our common-sense psychological intuitions, we accept this aspect of Martian memory into the cognitive fold, surely only skin-and-skull prejudice stops us from extending the same courtesy to Otto” (91). Rupert, Adams and Aizawa are committed to the claim that extended processes are functionally unlike internal cognitive processes and so do not deserve to be called cognitive. Their objections are that fine-grained features of human cognition are necessary for mentality. Yet Sprevak notes that this seems wrong, as the Martians

could differ in all kinds of fine-grained psychological ways and still have mental states. Sprevak notes that, with regard to the degree of how fine-grained one's functionalism is, "if this parameter is set too fine, then one is committed to Martians who differ from us in minor ways not having mental states. If the parameter is set too coarse, then functional role specifications are too easy to satisfy, and systems that are intuitively non-mental wrongly count as mental" (510). If we infer that the firing of c-fibres is the only way that pain is realized, then we leave out the very possibility that different species on earth with wildly different psychological structures can ever instantiate pain.

### 2.3. Sprevak and Miyazono

Sprevak's claim is that the functionalism that motivates the Martian intuition entails that HEC is true. "My claim is that if the grain parameter is set at least coarse enough to allow for intelligent Martians, then it also allows in many cases of extended cognition" (510). If a functional theory sets the level of grain too fine, then intelligent Martians are not allowed, and HEC is false. However, the Martian intuition pump is the sort that most would not be inclined to reject. This allows for HEC to be true.

Sprevak does some incredibly interesting philosophical work here. While his article responds to the criticisms of Rupert, Adams and Aizawa, he proposes a novel criticism against HEC. The extra conditions to the functionalist argument that Clark and Chalmers add make HEC a more modest theory. However, Sprevak claims that this modest form of HEC is incompatible with a functionalist defense of HEC. For instance, Clark and Chalmers claim that an external resource needs be reliably available and typically invoked (Clark and Chalmers 17). Yet it seems entirely possible that one can use the Martian intuition to imagine a Martian with internal cognitive resources that are neither reliably available nor typically invoked. "If the functional

roles of cognitive states and processes are specified broadly enough to allow for *internal* resources not to be reliably available or typically invoked, then they should allow *external* resources not to be reliably available or typically invoked either” (Sprevak 514, author’s emphasis). Sprevak claims that requiring constraints on the external but not the internal conflicts with the assumptions on which the argument for HEC was based. It is in this vein that Sprevak shows that the functionalism motivating the Martian intuition and HEC cannot be legitimately constrained.

Functionalism entails HEC, but not the modest version Clark and Chalmers posit. Sprevak claims that this radical HEC is almost certainly false. Sprevak gives the following example:

One could imagine a Martian whose memory, instead of being stored in patterns of neural activity, was stored internally as a series of ink marks. If the Martian wished to store new information, it would activate a process that would create new ink marks in its storage system. If it wished to retrieve information, it would activate a process that would make a mental image of ink marks appear in its visual system. It seems wrong to say that simply because a Martian stored its memories this way, we should deduce that it had no mental life, or lacked genuinely mental memory states or processes. In principle, there seems no bar to such a Martian having beliefs and mental states (provided, for example, it exhibited the relevant course-grained features of memory, and caused occurrent beliefs and desires in an appropriate way). (508)

From this example it follows that, granting the functionalist extension, simply by picking up a book one comes to believe everything contained in that book (517). He justifies this claim as follows: as well as acquiring beliefs via its senses, Sprevak wants to grant that it seems possible

for such a Martian to be born with innate beliefs. Furthermore, it seems possible for an organism to have innate beliefs that it has not examined yet—a library of data that is hard wired into the organism by developmental processes, which the organism has not yet had cause to employ. Imagine that an ink-mark-based Martian is born with a stock of innate beliefs, most of which it has not chosen, or had cause to examine yet, but it could if it wanted to. It seems conceivable that such a Martian could exist. The Martian has ink marks inside its head that, if it were sufficiently diligent, would guide its action in appropriate ways (Sprevak 517). The idea is that, when reading a book, one has ink marks outside one's head. If one is sufficiently diligent, these marks guide actions in a suitable manner. By the extension of the analogy, if the Martian has belief, then so does the person with the book.

I think one initial response to this scenario is to deny that the ink blots inside the Martians head constitute beliefs just by their presence. Rather it seems that one could require that such a Martian actively endorse and analyze such innate ink marks prior to acquiring them. If this is what Sprevak means when he requires that the Martian be sufficiently diligent for the markings to guide its behavior in appropriate ways, then it seems obvious that one does not come to believe everything contained in a book simply by picking it up. Rather, if the agent is sufficiently diligent, the markings on the book will guide the agent's actions in appropriate ways, most likely to either the conscious endorsement of the content of the book or the disregard for the content. The criticism here is that it seems that Sprevak is watering down the concept of belief. A similar response has been motivated by denying that the Martian has innate beliefs that are stored as ink-marks: its beliefs are not realized by virtue of the intrinsic storage of the ink-marks (Walter 2010; Wheeler 2010).

Ignoring these criticisms for now, it appears that radical HEC seems false and should be rejected in order to maintain our understanding of how common-sense mental events operate.

Motivating this conclusion is the fair-treatment principle. The fair-treatment principle is the notion that, “if the only significant difference between two processes is that one is extended and the other internal, then both should have an equal claim to mentality” (Sprevak 520). If we hope to save HEC we will need to salvage a more modest form of HEC by finding some other conditions that can be added without disrupting the functionalist argument. Sprevak believes that it is unlikely that such conditions will be found. For one, any conditions would have to satisfy not just actual extended cognitive systems but also by all actual and possible internal cognitive processes. “Given the vast variety of possible internal cognitive processes, such a condition would hardly be any constraint at all” (519). Secondly, Sprevak claims that it is unclear how adding an extra constraint would help to avoid radical HEC anyway, as adding an extra condition does not, by itself, disrupt the feasibility of the Martian scenarios. Sprevak claims that the only way to avoid radical HEC is either to drop the fair-treatment principle or to drop the claim that Martians in these cases have mental states. The first is “an unacceptable way to defend modest HEC” and the second allows the criticisms of Rupert, Adams and Aizawa to return (519).

One way of responding to this challenge by Sprevak is provided by Miyazono (2017). Miyazono does not believe that the functionalist argument motivating HEC results in the radical conclusion Sprevak claims it does. Instead, a Systems-Reply (SR) would be the superior way of conceptualizing the functionalist implications of the Martian intuition. He believes that what actually follows from Clark and Chalmers’ argument is not that the beliefs of Otto are physically realized in the notebook, “but rather that the beliefs of the hybrid system consisting of Otto and his notebook are physically realized in the notebook” (3531).

The SR is derived from the basic ideas of a response to the Chinese room argument. The Chinese room argument was made to refute strong AI through the following thought

experiment. A man, let's call him John, is placed in a room with a ledger that contains instructions in English for manipulating Chinese symbols. The room takes inputs (Chinese symbols) which John can utilize in conjunction with the ledger to output a different Chinese symbol. Let's also say he has a button, because that is the theme of this chapter apparently. To an outside observer, the responses are perfectly reasonable to the inputs. Strong AI claims that there is nothing more than implementing a right program for understanding, and that John is simply implementing the right program. However, the SR claims that it is the John-room system, rather than John, that is implementing the right program. Thus strong AI entails that it is not John but the John-room system that understands Chinese.

The SR to Sprevak has a similar structure to the SR to Searle. Remember that Sprevak articulates the implications of radical HEC through the idea that, through the acceptance of ink-blot memory Martians, one comes to believe the contents of a book simply by picking it up. According to Sprevak's notion of SR, "we see no relevant functional difference between the role of the book for the Mark-book system (rather than for Mark) and the role of the innate storage for the Martian" (3530). The functionalist argument entails that the Mark-book system's belief, rather than Mark's belief is physically realized in the book (where belief refers to the contents of the book that Sprevak claims one believes by virtue of picking up).

The running theme of the last section is of warring authors who are arguing against each other to show who has the best argument against Clark and Chalmers. In this same vein, while Miyazono has shown a way that the radical HEC thesis is not entailed by Clark and Chalmers' functionalist argument, he is going to show how the functionalist argument does not entail what Clark and Chalmers' claim. He claims that Clark and Chalmers unjustifiably identify the original cognitive subject with the hybrid one (the Mark-book system) in order to derive their conclusion about the extension of the mind. When applied to the example of Otto, Miyazono claims that he

sees no relevant functional difference, “between the role of the notebook for the Otto-notebook system (rather than for Otto) and the role of the internal memory storage for Inga” (3533).

One reason for thinking that there are relevant functional differences between the role of the notebook for Otto and the role of internal memory for Inga is in terms of access. Inga’s access to her internal memory is introspective, while Otto’s access to his notebook is perceptual. Authors have argued that the functional difference between perception and introspection is significant even from the functionalist point of view (see Gertler 2007). However, Miyazono claims that Clark and Chalmers theory responds by positing that while the access to the notebook is perceptual for Otto it is introspective for the Otto-notebook system.

I question the effectiveness of the above claim. The parity principle and the functionalist argument are the claims that, all things being equal, we should individuate functional roles in terms of commonsensical, coarse-grained descriptions rather than empirical, fine-grained descriptions. Thus, while perceptive and introspective access may have fine-grained differences, we can view them as sufficiently equivalent when differentiating via functional roles. Therefore, while the claim that the Otto-notebook system is more similar in a fine-grained sense to Inga in that both can be articulated as having introspective access, the functionalist argument itself is motivated by the notion that this distinction between perceptual and introspective access is irrelevant in that the two forms of access can be characterized by the same functionalist grounds.

Miyazono might respond here by saying that, regardless of the functionalist attitudes framing this discussion, the fine-grained distinctions are significant enough to motivate the view that Otto’s access to his notebook differs significantly from the Otto-notebook system’s form of

access. One thing he says is that, “it does not seem to be part of our commonsensical understanding of beliefs that beliefs are, or can be, accessed perceptually” (3534). Thus, we don’t run into conflicting ideas of what beliefs entail when we understand the hybrid-system as having access characterized by introspection. Miyazono claims that Clark and Chalmers support the idea that there is no relevant difference between the role of the notebook for the Otto-notebook system and the role of internal memory storage for Inga through noting that Clark himself says that, “from the extended point of view, Otto’s inner processes and the notebook constitute a single cognitive system. Relative to this system, the flow of information is wholly internal and functionally akin to introspection (Clark 2011, 100).

The above argued that the differences of access between Otto and his notebook and Inga are not functionally relevant. Another way to complicate Miyazono’s claim that Inga’s access is functionally similar to the Otto-notebook systems access, in that the two are both characterized, is by denying that the functional argument only goes so far: one can deny that the Otto-notebook system’s access is introspective based upon fine-grained differences between perception and introspection. Even if we would like to characterize the ways that a hybrid system cognizes as sufficiently analogous to introspection, there are still ways that this presentation can be complicated. For instance, the bare fact that these hybrid systems incorporate both introspection and perception into their forms of cognitive access is enough to say that the process cannot be fully characterized as introspective. Take a different strategy: we might think of our introspective access to our mental states as immediate and non-inferential, providing an unmediated and privileged form of access. Perception, on the other hand, can be characterized as being mediated through sensory organs and thus makes use of sensory perception in ways that introspection cannot (for instance, when we introspect, we do not smell mental states).

There is one way that I think we could potentially allow Miyazono a defense against the claim that the analogy between the hybrid system and Inga can still be defended. Higher order theories of consciousness argue that what makes a conscious state conscious is that the state is the intentional object of, or is represented by, another of the subject's mental states, suitably placed or employed (Lycan 94). Higher-Order Representationalist (HOR) theories of consciousness try to explain the distinctive properties of consciousness in terms of some relation obtaining between the conscious state in question and a higher-order representation of some sort. For a state to be a conscious state is for said state to be represented by another mental state. Inner sense theory, or higher-order perception (HOP), is a higher-order theory that claims that, "a mental state is conscious just in case it is the object of a kind of internal scanning or monitoring by a quasi-perceptual faculty" (Lycan 95). Thus, consciousness is the functioning of internal attention mechanisms directed upon lower-order psychological states and events.

Therefore, if we allow that consciousness is partially maintained by an inner-sense, a kind of quasi-perceptual monitoring faculty, then we have reason to think that introspection, insofar we consider conscious introspection, is like perception in significant ways. In this way, we can defend the idea that the 'introspection' of hybrid systems is sufficiently like Inga's introspection, in that both cognitive process can be characterized through perception<sup>7</sup>.

---

<sup>7</sup> Simultaneously, it should be noted that HOP theory is not the only theory that attempts to describe consciousness. Just within the vein of contemporary HOR theories there is significant objections to HOP theory. For instance, Carruthers (2000) notes that internal monitors postulated would need to have considerable computational complexity to generate requisite higher-order experiences. The mechanisms in question would need to be explained by a plausible account of evolutionary pressures that led to its construction. Yet this theory lacks a substantive account of what these mechanisms would be like besides pointing out that they function like instances of external-perception. Yet if they function like external-perception, should inner-sense mechanisms be subject to malfunction like external perception organs encounter? There are reasons to question whether HOP theory provides the best account for higher-order consciousness, let alone with the multitude of competing theories that claim to analyze consciousness correctly. However this thesis does not endeavor to discuss competing theories of consciousness. The discussion of theories of consciousness here was undertaken to illustrate the functional similarities between perception and cognition

It is not clear that the claim that introspection is more similar to perception strengthens Miyazono's case. Claiming that introspection and perception cannot be shown to have both coarse and fine-grained differences does not only provide support for the claim that the hybrid system and Inga are similar, but also provides defense for the claim that Otto and Inga are similar. For if we can think of Inga's introspection as being, in some ways, characteristically like perception, then the functionalist argument is only made stronger with the recognition of the fact that both can access memory in similarly characteristic ways. Since the connection between Otto and his notebook is defined by Otto's perceptual access of it, we can think of Otto and Inga being alike in that both of their forms of access can be characterized by perspective. The point of Miyazono's claim was that Otto-notebook system is more functionally similar to Inga than to Otto, because both can be characterized as having introspective access while Otto only has perceptual access. Yet if we allow that introspection can be characterized as functionally similar to perception, then we have no reason to think that Inga is more functionally similar to the Otto-notebook system than to Otto. The above conclusions may seem contradictory at times, as I have argued for multiple ways of evaluating the analogies between Inga, Otto, and the Otto-notebook system. The purpose of the above was not to propose a correct way of reading the analogy, but to show that Miyazono's initial analogy can be complicated in a multitude of ways.

Miyazono goes on to investigate three different objections, two of which are essentially knock down for him. The most interesting objection is the Otto-as-the-System Objection (OAS). The objection posits that Otto is identical to the hybrid system. Miyazono provides two responses to this, the first that the OAS objection might make the functionalist argument circular. "The problem is that if their justification for Otto-as-the-system comes from OEB (Otto's extended belief: Otto's belief that the museum is on 53rd street is physically realized in his notebook), then they cannot, without circularity, defend OEB on the basis of OAS. They

cannot, without being circular, argue that, on the one hand, Otto is identical with the hybrid system because Otto's beliefs are physically realized in the notebook and on the other hand, Otto's beliefs are physically realized in the notebook because Otto is identical with the hybrid system" (3536). This objection need not be as damning as one may think, even though Clark and Chalmers do not provide an independent argument for OAS. For even though they do not provide reasoning for why we might think of Otto as the system, an independent argument (or investigation) can still be conducted.

#### 2.4 Milojevic

One final complication that will be addressed is brought up by Milojevic (2018). Recall that Miyazono claims that Clark and Chalmers unjustifiably identify the original cognitive subject with the hybrid one in order to reach their conclusion about the extension of mind. Miyazono's SR posits that it is hybrid cognitive systems that are the subjects of extended mental states. This changes HEC, and subsequently both extended mind and extended self, from the radical thesis it is. "It would become a much less interesting and radical thesis, more similar to the hypothesis of strong AI, stating that there could be systems partly organic and inorganic whose states can play relevant causal roles" (Milojevic 9); "Indeed, what makes the idea of extended mind original is the claim that it is Otto who believes that the museum is on 53<sup>rd</sup> street. The originality is lost if the claim is just that the Otto-notebook system believes that [the] museum is on 53<sup>rd</sup> street" (Miyazono 3539).

Miyazono identifies one possible avenue of response to his SR. This is through the OAS objection, the claim that Otto is identical to the hybrid Otto-notebook system. In regard to the OAS objection, Miyazono claims that the functionalist argument cannot justify the assumption about the identity of these two kinds of subjects without circularity (See 2.3). Thus, the

dilemma: abandon OAS and make EM an uninteresting thesis, or endorse it and make the argument for EM circular. A third option is to offer independent justification for OAS. Milojevic attempts to provide this independent justification through the following strategy. By allowing that there are hybrid integrated systems (i.e., Mark-book and Otto-notebook systems), Milojevic provides an argument that sometimes these systems have mental properties. From there, the claim is made that these systems are, at least sometimes, identical to subjects that are extended by processes of hybridization (10).

The first objective Milojevic pursues is showing that there are hybrid integrated systems. Milojevic posits that, in order to call a process or mental state cognitive, it has to be a part of an integrated system with appropriate properties. She also posits that an underlying system that enables *causal interactions* between mental states and cognitive processes yields intelligent behavior and agency. Thus, showing that relevant systems are appropriately integrated is a crucial step in her argument; that these systems are capable of having mental states. Milojevic claims that we must offer an independent account for integration and for what makes a part of the world a part of a cognitive system, as the investigation of hybrids and their states, as it is done in the research of the extended mind thesis and HEC, cannot bootstrap entirely off of the systemic organization of biological organisms (Milojevic 11).

In understanding the question of what makes a part of the world a part of a cognitive system, Milojevic states that, "integrated cognitive systems are integrated physical systems, meaning that their parts are mutually coupled and exhibit dynamical and causal regularities in their interactions, but they also exhibit typical forms of cognitive integration in terms of the cognitive functions that different kinds of flows and transformations of information perform" (12). The idea is that it is a specific function that different parts of the world perform that makes them a part of the same system. While some systems will have, say, mental states which will be

interconnected in a systemic way reflecting specific psychological functions and regularities, it is important to note that conditions of particular cognitive integration can vary. Integration can occur in a number of ways depending on the contingent circumstances of a particular cognitive task or function (12). For instance, with the case of Otto and his notebook, the integration is explained by the fact that the information in the notebook was previously consciously endorsed, it is readily available, is more-or less automatically endorsed after retrieval, and is easily accessible. Thus, it is not enough to point out that specific conditions of integration are not met in some cases in order to prove that they are not conditions of integration at all. If one accepts this integration assumption, then one does not need to dispense with the conditions used to support a modest HEC.

With the above argumentation in place, Milojevic believes that she has shown that there are hybrid integrated systems. Not only this, but this claim for integration allows us to infer a substantive difference between the Otto-notebook case and the Mark-book case. Mark and the book are not sufficiently integrated and do not satisfy the inferred conditions (i.e the glue and trust). Before moving on, I would like to readdress something. The reason that Sprevak talks about the Mark-Book system in the first place is as an example of the absurdity of some instances of HEC when the conditions that would constrain it are lost. Sprevak argues that we cannot legitimately constrain functionalism as it will not allow us to conceive of Martian cases. If we agree with Milojevic and necessitate these conditions as minimal integration conditions, are we committed to non-intelligent Martians? Initially, I don't think that we need be pushed this far. These minimal integration conditions for Otto and his notebook can vary from the integration conditions for other cognitive systems. "Integration can occur in a number of ways depending on the contingent circumstances of a particular cognitive task or a function" (12). One thing that Sprevak complicates are the conditions of easy accessibility and ready

availability. He introduces the possibility of a Martian that has a cognitive process that is not easily accessible (and likewise that there are some human cognitive processes that have corollaries, like memories, that are not frequently or easily accessible). By the following logic of the condition, the Martian is not exhibiting cognition. Thus, the condition of easy accessibility/ready availability limits the functionalist argument and, by the principle of fair treatment, should be dismissed. Yet, if we allow Milojevic's consideration, we need not think that the discussed principles explain the integration conditions of all cognitive systems<sup>8</sup>, and thus that the characterization of the integration of a Martian's cognitive systems need not be explained by the same conditions that explain the integration of Otto and his notebook.

Next, Milojevic briefly addresses the second step: providing an argument that sometimes hybrid integrated systems have mental properties. She does not offer much detail, simply stating that this step explicitly employs the parity principle, which should guide the identification of mental states of systems by referring to common sense functional roles. The argument that hybrid systems have mental properties is for Milojevic analogous to the argument that mental states extend beyond the body. Important to note, the first step is necessary in the assertion for the conditions (glue and trust) that allow us a more modest HEC. Therefore, we can see the two steps as contingent upon each other.

With all of this in place, Milojevic believes that she can provide a response to the SR against HEC. This is done by claiming that Mark and the book are not sufficiently integrated and

---

<sup>8</sup> This being said, it seems that each novel instantiation of some identifiable part of a cognitive system will need an account for why it is integrated in such ways within the system. While some backpacking can be done (for instance, inferring that cognitive process x exhibit the same typical forms of cognitive integration in terms of the cognitive functions that different kinds of flows and transformations of information are performed in cognitive process y, then it is most likely that the integration conditions for underlying parts in the two processes are similar), it seems that most of the time a new account will be necessary.

thus that the Mark-book system does not have any extended mental states. This means that, “we are not offered good reasons to believe that Otto is different from the Otto-notebook system in the first place” (15). The initial claim by Miyazono was that we should differentiate between Otto and the Otto-notebook system, and that the extended beliefs should be attributed to the hybrid system rather than Otto. This is because of the analogy between Mark and the Mark-book system that was established on differential attributability of beliefs as some beliefs are attributable to the system but not to Mark alone. However, if we don’t have good reason to believe that the Mark-book system has mental properties, then we don’t have good reason to believe that Otto is different from the Otto-notebook system (15).

Milojevic leads the discussion into one of how the OAS claim can be made sense of through personal identity. While we don’t have good reason to believe that Otto is different from the Otto-notebook system, we still need some kind of account for why we might think of them as similar. Milojevic believes that OAS, when it claims identity between Otto and the Otto-notebook system, cannot be a claim of physical identity. Rather, it can be sensibly interpreted as, “a claim about the numerical diachronic identity of a person [the claim that the-biological-Otto-at-T1 is identical to the-Otto-notebook-system-at-T2]” (16). Thus comes about the discussion of EST: if personal identity is sustained through the process of hybridization then mental states, “which are realized in both biological and artifactual parts of the system, are *attributable to the person that entered the process of hybridization and thus, constitute a part of the extended mind*”<sup>9</sup> (16, emphasis added).

---

<sup>9</sup> Just as a clarification here: the purpose of this first chapter is not to provide a systematic definition of either EST, personal identity or personhood. Much of that discussion will be done later in discussing how I believe we can make sense of EST through the philosophical literature surrounding personal identity and personhood. However, Milojevic’s articulation of personal identity is part of her defense of the extended mind thesis, and thus in adequately representing how the literature has responded to HEC, illustrating all of her defense against Miyazono is required here. This being said, personal identity and personhood will

Milojevic makes reference to one view on the persistence conditions of personal identity, this view being the psychological continuity view<sup>10</sup>. This view essentially prizes the continuity of thoughts and memories of past experiences as what matters for personal identity. “A person is a changing entity whose identity is determined only by her psychological continuity, independently from her personal traits and specific parts of her body” (Milojevic 17). One aspect of the theory that Milojevic emphasizes is that a person can gain or lose qualities and capacities and still be characterized as the same person as long as sufficient psychological continuity is held. The larger claim here is that, if we accept the psychological criterion as accurately describing the nature of persons and personal identity, and if we can aptly recognize the hybrid extended systems as psychologically continuous with Otto, then the hybrid Otto-notebook system is identical to Otto and is thus a constitutive part of Otto.

Thus, Milojevic believes that not only is Otto still the same person after hybridization, but also that, “we can even claim that the retention of his personal identity was enabled by extension” (17). Importantly, even though we can consider Otto as a confirmation of the retention of personal identity through hybridization, Otto’s confirmation is only contingent; we can imagine cases in which there is no such retention. Think back to the original SR to the John-room system. The person inside the room, the leger, the database, and the important button constitute a system. By the functionalist view of integration these parts are sufficiently integrated and form a system, and that this system could not perform its main function without any of these component parts (Milojevic 17). By the SR, we can identify mental states of this system, but these mental states are not attributable to the person in the room. Thus, the

---

be defined in detail later, and thus the following discussion will only precede on the basis of how she uses the following terminology.

<sup>10</sup> This view will be analyzed and explicated in greater detail later. For those interested in better understanding the view, see Locke (1689), Parfit (1984), Shoemaker (2008).

inclusion of the criterion of psychological continuity as one of the criteria motivating Milojevic's analysis allows for a clear dissimilarity between Otto and John. "Narrow Otto and hybrid Otto are psychologically continuous because there are chains of psychological connections between their mental states, [while] the person in the Chinese room and the whole system are strongly discontinuous in a psychological sense. Otto's occurrent beliefs produced all of the Otto-notebook system's dispositional beliefs, wishes and desires of narrow Otto still guide the behavior of hybrid Otto, hybrid Otto remembers some of the experiences of narrow Otto, etc." (18). While we can identify beliefs of the system by the answers to the questions posed in Chinese (17), these answers and corresponding beliefs will not be appropriately psychologically connected with the mental states of John. Therefore, we can think of the psychological continuity criterion as providing the independent argumentation necessary for verifying OAS.

In closing this discussion of Milojevic's contribution to the discussion of HEC, I would like to note the following: Milojevic's discussion and defense of the extended mind argument was necessary to discuss here so as to adequately show how the literature has responded and adapted to discussions of HEC. While an exposition of her work was given here, we are not done with discussing her thesis. Time will be spent in the fifth chapter discussing her work.

### 3. Conclusion

In concluding this chapter, I would like to address a concern of mine. The first is that part of the project that HEC theorists concern themselves with is showing how atypical instances of cognition can be conceived as being extended into the environment. One serious premise in this kind of project is that the human mind is the 'taxonomical dictionary' we use to understand and identify some causal process as cognitive. This means that what we understand the cognitive to be will necessarily have to have some correlation with actual 'observable' processes

in the organism. This is for good reason, as it keeps HEC theorists in check when we consider what is or isn't a cognitive state. We avoid over attribution of cognition to the point that it bleeds out into every aspect of the world, effectively reducing the importance and philosophical force of HEC. This being said, this makes the kind type 'cognition' a kind that is entirely contingent upon contemporary cognitive science. It is a kind whose membership will oscillate as we identify further ways humans cognize and reject other previously endorsed views that posit a false notion of cognition. The notion of cognition, as we must understand it within the debate, can necessarily only work upon definitions of human cognition and thus is entirely dependent upon evolving scientific discourse on what exactly identifies the 'mark of the cognitive'. A dilemma thus emerges. The way that we define cognition will be incredibly important in future discussions of HEC, for we may develop the technology to implement profound changes in human cognitive processes. Say we imagine a human that is experiencing analogous P-conscious states as others through the routing and replications of these processes underlying consciousness. This possibility may or may not become an actual possibility in the future. Does this conceivable notion change our perspective of what counts as cognitive, or how it can be realized?

I am not so sure of this conclusion at the moment. However, in some ways, I am convinced by certain arguments put forward against HEC. I believe that the Martian example can only go so far. Likewise, regarding the earlier concession that we should be critical of arguments for functionally analogous equivocation, it seems hard to conceive of working memory as the sort of thing that can be properly conceived of as extended, as part of the conception of it is its limitations to a very small quantity of information. If we think of working memory as the capacity to temporarily store and manipulate information in the service of ongoing tasks (Baddeley 1986), or as, "the storage system responsible for the maintenance of

information in the service of ongoing work—that is, the system that makes available stored information for task-based information—without imposing a limit on its duration or relationship to long term memory” (Persuh et Al. 2018). An emphasized characteristic of working memory is its storage-capacity limit, and much work has focused on this aspect. Two studies (Miller 1956, Cowan 2001) reported an average capacity of seven and four items for verbal memory. This limited capacity-storage feature does not seem to hold intuitively once we allow for extension. If we were to say that one could expand their working memory through a notebook, then it seems that the system of storing these variables contained in working memory varies not only in fine grained differences (i.e one accesses these stored variables by introspection vs perception) but also varies in that the utilization of the notebook can expand the normal limits of working memories storage-capacity. Additionally, when the conception of the limited capacity of working memory is challenged, it seems that we lose an important criterion by which we understand working memory in the first place. In this regard, it seems apt to agree with Rupert in his argument against Rowlands when he says that working memory is better conceived of as internal than as a conglomeration of both internal and external stores plus the processes that operate on these stores (Rupert 409).

The second concern is of the relationship between HEC and EST. To note at the outset of this discussion: the preceding arguments largely represent objections to HEC and the arguments for the extended mind. It should be noted that this discussion will be supplemental to the later discussions of EST. While these objections against Clark and Chalmers are significant, I do not believe that the perceived success of these objections significantly detracts the forthcoming discussion of the tenability of the extended-self. “Cogent criticisms of HEC will not, of course, refute the hypothesis of an extended mind; given, however, that current work on extended cognition promises to provide the strongest support to date for the view that the mind is

extended, HEC's problems are, in no small measure, problems for proponents of extended minds" (Rupert 391). The articulation here is of how criticisms against HEC do not necessarily defeat the hypothesis for the extended mind. In the same way I would like to clarify the conclusion that these objections are strong enough to defeat HEC need not necessarily take away from any discussion of the tenability of EST. However, as Rupert notes, the acceptance of the bare possibility of HEC does nothing significantly interesting. In fact, he notes that without a positive affirmation or a, "promising framework for the pursuit of cognitive science (as it attempts to understand actual mental states), the radical theses of extended mind and extended self lose much of their current appeal; one cannot infer that the human mind or self is extended—or that we are creatures of the extended world—from a premise asserting the bare possibility of extended cognition" (393). This claim is significant, considering the undertaking of my endeavor. However, I believe that this chapter has provided a solid grounding and defense for HEC, and as such do not believe we need to worry about foundational difficulties in discussing the extended mind or EST seriously. For the time being, discussions of the plausibility of HEC will be partly put to bed. Obviously, some important themes will reemerge when addressing EST, but for the sake of a systematic investigation the following chapters will not address the tenability of HEC.

The next chapter will focus on discussing EST. This will be done through expounding the original claims by Clark and Chalmers for this form of extension, as well as articulating an approach to this analysis through the lens of personal identity and personhood. Following this will be a discussion of a few authors who reject the possibility of EST. The argument will be made that these authors hastily reject EST, and as such that EST is still a tenable thesis even in the face of criticism.

### Chapter 3. The Extended Self Thesis and Objections

#### 1. Introduction

In the previous chapter, both a review of the literature and a modest defense of HEC were provided. If anything is deduced from the last chapter, it is that HEC can still be thought of as a tenable thesis. The chapter ended with a discussion of Milejovic's defense of HEC through the claim that hybrid systems can be thought of as, under proper circumstances, involving the systematic integration of the relevant parts, identical to their original cognitive subject: the Otto-notebook system is identical with Otto himself and can be thought of as a constituent part of Otto. While a discussion of the extended-self was partially alluded to, it did not go into depth, as a thorough analysis of the extended-self thesis (EST) had yet to be carried out. The chapter closed by noting some issues surrounding the discussion and defense of HEC, but largely concluded with the assumption that HEC is defensible and can still be thought of as a tenable thesis. Thus, this chapter will proceed with the following assumption: HEC is a tenable thesis, and we can assume it as a valid premise in following discussions of EST.

The purpose of this chapter is to get clear on what exactly EST is. What exactly is claimed when Clark and Chalmers claim that we can see ourselves as extended selves? What will follow will be an investigation of EST through analyzing the ways that various authors have conceived of this thesis. Specifically, I will focus on authors who seek to show that EST is an unjustified and false thesis. However, prior to discussing their concerns with the thesis it will be necessary to illustrate how these authors conceive of this thesis.

This chapter will proceed as follows. I will start with a small preliminary discussion of personal identity and personhood. Much of the literature that responds to EST makes explicit reference to theories of personal identity and personhood, and thus a terminological grounding

of these concepts will help in grounding these discussions. Following this will be a discussion of the initial articulation of EST by Clark and Chalmers. After this will be a supplemental discussion of how other authors have endeavored to understand EST. Following this detailed exposition of EST, an analysis of the objections various authors have levied against EST will be undertaken, focusing explicitly on Baker (2009; 2013) and Olson (2011). The purpose of discussing these objections is not to provide a knock-down defense of EST, but rather to show, like the last chapter, that EST can still be conceived of as a valid and tenable thesis.

One thing to emphasize initially: this thesis is concerned with the tenability of EST. Part of the discussion of its tenability will be an analysis of EST and its implications. This thesis will follow the methodology of a substantial amount of the literature in assuming that EST is a thesis well suited to be evaluated through personal identity or personhood. When questioning whether Otto's notebook is part of him, or whether we can conceive of hybrid-systems as identical to their original cognitive counterparts, we are asking questions about personal identity conditions. We are questioning whether the skin-and-skull boundary is limiting in our ability to conceive of what persons are. Likewise, it should be noted that a distinction will be made between the philosophical conception of self and the concept of a person. This thesis is ultimately concerned with a discussion of the extension of persons and personal identity. Thus while there remains a distinction between the two, this distinction should not heavily impact upon this thesis' ability to address and discuss authors that reference either selves or persons in their discussion of extension. As will be addressed shortly, there are specific concepts and implications drawn when referring to an analysis of personal identity or personhood, as well as larger themes widely accepted by authors who talk about these concepts. Whereas this type of consensus is not found as generally when looking at philosophical conceptions of selfhood. Thus

while there is a distinction between selves and persons, this thesis does not endeavor to survey and explain the commonalities of different philosophical accounts of selfhood.

## 2. Preliminaries

Generally, two questions typically arise in the literature on personal identity: what is the nature of persons, and what criteria identify a person over time. Some authors propose that in answering the second question one is presupposing an answer to the first question: we need to know what persons are in order to identify them over time (Wilson and Lenart 424); “only if we know the conditions of personhood, can we give a compelling account of personal identity over time” (Wagner and Northoff 1). In answering the question of persistence conditions, one is concerned with diachronic aspects of personal identity, whereas answering the question of what a person is is to answer the question concerned with synchronic aspects.

Likewise, having an idea of how one answers the persistence question typically allows us to analyze synchronic dimensions of persons. For instance if we characterize F as a quintessential aspect of x, something that is a necessary feature of our identification of x as x, then it is possible that, in the framework of personal identity and personhood, F plays some kind of role in the attribution and articulation of the persistence of x. If we were to consider F as the retention and future utilization of specific individual memories, then we could say that Shelton can be identified as Shelton through these memories (and above this that memories as a general category are important for our identification of Shelton as a person).

Bear in mind however that how one answers the persistence question does not always allow us to analyze synchronic dimensions. It can be entirely possible that a feature that helps us answer persistence questions does not help us answer synchronic questions. Maybe Shelton is characterized by his neurotic obsession with buttons, and with the pushing of these buttons.

This aspect may be necessary in identifying Shelton as Shelton over time, but it does not inform us of the synchronic dimensions of personhood. (I will go on record saying that many people I know do not care for buttons and the pushing of them, and yet they can still be characterized as persons).

As such, theories of personal identity must at least implicitly presuppose a view of personhood, and accounts of personhood must at least implicitly consider how personal identity is constituted (Wagner and Northoff). This being said, one can ask questions of personhood without necessarily discussing personal identity in depth, and vice versa. Consider the following: an individual wakes up one day with some form of intense amnesia, forgetting all personal facts about their life. Grant also that this person retains basic motor skills, proficiency with language and sufficient self-awareness. We can analyze this case through both the lens of personhood and personal identity. For instance, we may infer that this person is still able to think morally, albeit that their moral framework is not informed by their lost personal history. If we had an analysis of personhood that posited that  $x$  is a person if and only if  $x$  is a moral agent and is self-aware, then we could reasonably conclude that this individual is a person. The question of personal identity is a different question in this case, and depending on one's analysis of personal identity, can be incongruous with our answer to the question of personhood. Grant some proto-Lockean conception of personal identity where psychological continuity via memory is what entails identity:  $x$  at  $t_1$  and  $y$  at  $t_2$  are the same if and only if  $x$  is sufficiently psychologically continuous with  $y$ . Let us call the individual who falls asleep with all of their memories intact  $x$  and the individual who wakes up absent this psychological continuity  $y$ . By our analysis of personal identity, we have reason to think that  $x$  and  $y$  are not the same person: they do not share personal identity. This being said, we can think of  $x$  and  $y$  as being persons, regardless of which persons they are.

One way to conceive of personhood and personal identity as distinct is through understanding what sorts of questions each raise. The concept of personhood raises questions about what conditions are necessary for an entity to be a person at a discrete point in time (Wagner and Northoff 2014). Analyses of personhood provide answers to the questions of what a person is, what are the necessary and sufficient conditions for something to count as a person, what ensures the continued existence of a person as a person, what can we consider people to fundamentally be, etc. (Olson 2015). As stated earlier, questions of personhood can generally be understood as characterizing the synchronic dimensions of a person's life.

Alternatively, questions of personal identity are questions of diachronic identity. Analyses of personal identity likewise attempt to provide answers to the questions of what makes x at t1 identical to y at t2, what changes entail that x would no longer be y if undergone, what kind of criterion establishes the relation between x and y that makes them the same person, etc.

Debates around personal identity are concerned with numerical identity, a concept that can be contrasted with qualitative identity. Qualitative identity can be a one-to-one relation, or a one-to-many relation depending on how coarse or fine-grained one defines the qualities grounding the relation. Numerical identity is a one-to-one relation. In questions about numerical identity, "we look at two names or descriptions and ask whether these refer to one and the same person at different times, or rather to different persons" (Wagner and Northoff 4). Take the following example. If I take two billiard balls and paint them black, then one can say that these two are qualitatively identical. Imagine one of these balls was white before I painted it.

With regard to numerical identity, these two now black balls are not numerically identical, but rather the previously white ball and the now painted black ball are numerically identical<sup>11</sup>.

Here is where it becomes tricky with regard to personal identity, as one needs to specify what exactly someone is referring to when they claim numerical identity between x and y. If they claim numerical identity with regard to shared personal identity, then it is whatever the analysis of personal identity posits (e.g., memories) that guarantees the numerical identity between x and y. Remember the previous example of the amnesiac man. Since personal identity is understood as psychological continuity across memories, the identity condition is lost between x and y. This does not mean that x and y do not share different identities between them. For instance, both x and y have the same body. Thus x's body is numerically identical with y's body, even though x and y themselves are distinct. While personal identity is concerned with numerical identity, numerical identity of some mental or bodily aspects of a person does not itself entail numerical identity of the sort that is relevant to personal identity.

The question of numerical identity largely may not apply to the question of personhood as a broad analysis. Not only is numerical identity important for personal identity in that personal identity is concerned with persistence, but also in that personal identity is generally conceived as a one-to-one relation. A person cannot be in two places at once. More specifically, there cannot be more than one instantiation of the same person, where both individuals share

---

<sup>11</sup> Numerical Identity will become an important issue in the fifth chapter when we talk about how we can feasibly consider personal identity as applied to extended persons, and also in talking of Derek Parfit's view. Parfit believes that certain important questions presuppose a question about personal identity: there is a belief that the nature of our own identity through time is, "such as to guarantee that in these [problem] cases questions about their identity must have answers" (Parfit 3). It is through his example of a case of brain fission that the question of personal identity is made distinct from questions of survival. It is in this notion, the notion that we can evaluate whether a psychological entity survived but cannot presume questions about personal identity, that we may find a complication for the EST. Necessary for this discussion will be an understanding of the notion of numerical identity, and how we can use this to make sense of personal identity.

personal identity but occupy different spatial locations. This notion of numerical identity will be investigated in more detail later, as it will be a very important question whether extended mental states can be considered numerically identical with the original cognitive agent or person. But for now we will simply go forward keeping in mind that numerical identity will be important for our discussions of the extension of personal identity.

The above was a short exposition of the terms 'personal identity' and 'personhood' for the purposes of adequately understanding the coming discussions of authors who challenge EST and the extended mind thesis. More detail and examinations of theories of personal identity and personhood will be given in the fourth and fifth chapter. For the time being, we will accept the above definitions as adequate for our purposes and move on.

### 3. The Extended Self

The previous section saw the introduction of the framework of personal identity and personhood. The reason for this is that, as we will find out, most authors who seriously consider the possibility of an extended-self consider it from within the framework of personal identity and personhood.

In nearly every significant paper that deals with the notion of the extended self, authors makes sure to include the block quote by Clark and Chalmers that articulated this notion. As it is this quote that is the larger focus of this thesis, let us acknowledge the claim once more.

Does the extended mind imply an extended self? It seems so. Most of us already accept that the self outstrips the boundaries of consciousness; my dispositional beliefs, for example, constitute in some deep sense part of who I am. If so, then these boundaries may also fall beyond the skin. The information in Otto's notebook, for example, is a central part of his identity as a cognitive agent. What this comes to is that Otto *himself* is

best regarded as an extended system, a coupling of biological organism and external resources. To consistently resist this conclusion, we would have to shrink the self into a mere bundle of occurrent states, severely threatening its deep psychological continuity. Far better to take the broader view, and see agents themselves as spread into the world. (Clark and Chalmers 18)

There are several aspects of this passage to flesh out. First, it is largely unclear on the basis of this text alone what a 'self' is, according to Clark and Chalmers. Here is what we do know of the 'self' the authors talk about: it outstrips the boundaries of consciousness, so the self is partially constituted by consciousness but not entirely, or the self is something that supervenes upon consciousness but is not reducible to it.

At another point Clark and Chalmers note that Otto's notebook is a central part of his identity *as* a cognitive agent. They follow this by saying that the best course available is to see Otto himself as an extended system. This implies that one's identity as a cognitive agent is important in how we view the self that is Otto, granting that when we say that Otto is an extended system, we are referring to the self as an extended system. Thus Otto's identity as a cognitive agent bears on his self. In fact, by the claim that we ought to, "see agents as spread into the world" (18), it can be argued that Otto's identity as a cognitive agent is part of the identity of his self. Let us at least think for now that being a cognitive agent is a necessary condition for being a self (if not the stronger claim that one's identity as a cognitive agent is one's identity as a self). For the time being it would be wise to consider the former statement as the more accurate one, given that the self can be considered to outstrip the boundaries of consciousness. Unfortunately Clark and Chalmers do not give us an understanding of what else the self might outstrip. Perhaps what they are saying is that the self can be characterized by

unconscious processes? Or perhaps the idea is that there are non-cognitive aspects that can be deemed as constitutive of the self?

Importantly, it is claimed that Otto's notebook is a central part of Otto's identity as a cognitive agent, and with this it is best to regard Otto himself as an extended system; a coupling of biological organism and external resources. This has two implications for our conception of self. If something is a central part of one's cognitive identity, we can consider it largely part of the identity of the shared identity of the cognitive system. Secondly, by stating that we can think of Otto himself as extended, the notion is that it is the notebook that becomes (or at least is recognized as) part of Otto. It is unclear from this whether the authors mean to articulate that it is the extended system that can be identified as identical to Otto, or whether in a more literal sense that it is the notebook that is Otto.

Finally, resisting Clark and Chalmers' conclusion and thinking of the self as not extended is to 'shrink' the self to a mere bundle of occurrent states. This, they argue, threatens the deep psychological continuity of the self. This tells us a few things. It seems that the self can be characterized, at least partially, by psychological continuity, as inferred by the claim that shrinking the self jeopardizes the psychological continuity of the self. Granting a Lockean approach, if psychological continuity is a necessary condition for a self, then we can see extension as a process that enables psychological continuity.

With all of this said, it does not seem that we have a clear conception of what a self is. This is problematic, considering the competing analyses of selfhood. Take Dennett's (1986) conception of the self for example. A centre of gravity, as it is conceived of by Dennett, has no physical properties except for spatio-temporal location. It is a fiction, an abstract object, even though it has a well delineated and well behaved role within physics. Thus while a centre of

gravity is a purely abstract object, “it has a spatio-temporal career, which I can affect by my actions. A centre of gravity is just an abstractum. Its just a fictional object[...] that has a perfectly legitimate place within serious, sober, *echt* physical science” (Dennett 2).

Dennett wants to claim that the self, like a centre of gravity, is a theorist’s fiction. The self is like that of a fictional character (think of Sherlock Holmes or Jon Snow). Granted there are differences between these two types of fiction, many ways of understanding fictional characters aid in our understanding of the self. Dennett posits that fictional characters are usually encountered *fait accompli* (6): once the book or novel has been published it is too late for the novelist to render determinate or indeterminate any other details. When fans ask an author about more details about a story they have already written, they are asking the author to take the indeterminate fiction they have established and enlarge it. Dennett posits that this process is akin to asking an author to write a new novel, or to invent more novels for us, on demand. In contrast to selves, Dennett proposes that this process is a familiar process in that this is the way we treat each other. “Our selves are constantly being made more determinate as we go along in response to the way the world impinges on us. It is possible for a person to engage in auto-hermeneutics, interpretation of one’s self, and in particular to go back and think about one’s past, and one’s memories, and to rethink and rewrite them” (Dennett 7). This process does change the “fictional” character, the character that you are, in much the same way as an author makes more determinate his fiction through the addition of another novel. Dennett posits that this process of the determination of aspects of one’s self would be, “an utterly mysterious and magical prospect (and hence something no one should take seriously) *if the self were anything but an abstractum*<sup>12</sup>” (Dennett 7, authors emphasis). Like fictional characters, our own personal

---

<sup>12</sup> Granting Dennett’s claim that fictional characters and selves undergo the process of determination in a similar manner, it is not necessarily clear that one is forced into the conclusion that this process only

narratives can have contradictory properties. Rather than explaining this in selves through notions of cognitive dissonance, Dennett explains this by saying that we are all, “at times, confabulators, telling and retelling ourselves the story of our own lives, with scant attention to the question of truth” (Dennett 8).

Dennett’s larger claim is that the chief fictional character at the centre of one’s autobiography is one’s self. “When one asks the question of what the self *really* is, one is simply making a category mistake” (Dennett 10). This conception of selfhood is not necessarily at odds with Clark and Chalmers’ conception of a self. Dennett could agree with Clark and Chalmers that by narrowing the self to the confines of the body we would not adequately capture all of the fictional apparatus and externals that contribute to the confabulation of the self. However, if we grant Dennett’s conception of the self, as well as the notion that this conception of self is compatible with Clark and Chalmers’ hypothesis of the extended mind, then we are left with a variant of the uninteresting thesis talked about by Miyazono. Sure, the self can be thought of as extended. But this extension is of a different category. The initial thesis is a problematic and interesting one, suggesting that we ought to conceive of individuals, of selves, as extended into the environment. This thesis of the self extending is of a fictional entity being conceived of as part of the environment. Otto’s notebook is part of him as much as we conceive of the fictional entity as being conceived in part with the notebook. There is no actual substance beyond an established fiction that is extending.

There are a few avenues available in response to this situation. The first is to allow that Clark and Chalmers conceive of the self in the same way Dennett conceives of the self, and thus

---

makes sense if the self is an abstractum. Narrative theories of personal identity (see Schectman 2014 for example), allow that personal identity is constituted and in fact is guaranteed by this narrative process of self-identifying and analyzing different constitutive determinate aspects.

that the radical thesis about the extension of the self is not in fact such, and is more of a secondary implication for the way we utilize the language of selves and agents. It will have an impact upon the language we use to describe selves, but will not change the reality of selves as selves are a mere fiction. I do not believe this option to be viable. For it seems that Clark and Chalmers want to posit that Otto is a coupling of biological and inorganic parts: that Otto's notebook is *literally* a constitutive part of Otto. This conception of Otto's self as being such a way is problematized with Dennett's analysis, as Dennett says that questioning what the self is *itself* a category mistake. The second avenue is to say that perhaps Clark and Chalmers had a different conception of selfhood in mind, and thus that their thesis is still radical, if mysterious, due to the unspecified and unquantified nature of what an extended self would amount to. The third avenue is, realistically, that Clark and Chalmers did not have a rigorous conception of selfhood in mind when they made their passing comment about the possibility of the self extending. I am inclined for the time being to think that the last avenue is perhaps the most realistic one.

Thus I propose the following: Let us go forth and investigate how other authors have conceptualized EST. Through this investigation we can garner a better idea of what exactly it is that EST amounts to. Ultimately, it will be shown that many analyses of EST propose understanding it within the language of personhood and personal identity rather than selfhood. Regardless of which framework EST ought to be perceived through, this thesis endeavors to investigate it through the framework of personhood and personal identity.

What follows is an investigation of how a few prominent authors have conceptualized EST. Following this will be a specific analysis of various authors who reject EST. The analyses that attempt to problematize EST will not be the only ones investigated in this thesis however. But for the sake of orientation I will analyze a few of these authors that explicitly reject EST in this

chapter. There will be points of reiteration when explaining and analyzing the views advanced by Baker and Olson. While steps have been taken to avoid undue reiteration, there is a point in which aspects of their views will need to be explicated in both understanding their analyses in relation to personal identity and personhood, as well as evaluating their arguments.

#### 4.1 Contextualization: Olson

The first author who investigates the feasibility of EST is Eric Olson (2011). Ultimately, Olson will conclude that EST is largely unattractive and only follows from the extended mind if one thinks of 'beings' as literal bundles of mental states (481). First Olson endeavors to clarify that his analysis of EST is contingent upon a solid understanding of extended mind and what that entails. "What the extended self amounts to, and whether it follows from the extended mind, depends in part on what the extended mind amounts to" (482). Through reference to Clark and Chalmers, Olson notes that the extended mind thesis is that mental states and processes can be 'extended' or 'external' in the way that some of Otto's memories are supposed to be, i.e., the notebook containing the memory that the MOMA is at 53<sup>rd</sup> St. Thus Olson conceives of external mental states as those that extend beyond the brain or body. Olson notes that we need not take these states to be literally located beyond the body. "Even if mental states have no location at all, things located beyond the skin may be 'directly involved' in them, forming part of their physical basis or realization. So we can think of external mental states as those at least partly located or realized beyond their subject's skin" (483). Following from the claim that "nothing can literally be located in a place where it has no parts" (484), Olson proposes three ways of understanding EST: 1. There could be a psychological being that extended beyond its skin. 2. It is possible for a human person to extend beyond her skin. 3. Any psychological being could extend beyond its skin. Olson goes with the third possibility as it allows the strongest way of conceiving of the extended self (and, as will be shown in the analysis

of Olson, by refuting the strongest notion he thereby also refutes the two other possibilities). Thus granting that Otto is a psychological being, then if Otto extends beyond his skin because some of his beliefs are located in the notebook, then the notebook is a part of him

Olson notes that the extended self has an important mereological corollary, that being that, "Clark and Chalmers' remark that Otto would be 'a coupling of biological organism and external resources' suggests that he would have both the organism and the notebook as parts: he would be a disconnected or 'scattered' object" (484). Thus by a psychological being extending we can think of the psychological being as having parts located beyond its skin. In regards to the question of whether the extended self follows from the extended mind, Olson notes that it can only be because a psychological being must extend as far as its mental states extend. Thus Olson will argue that the link between extended mind and extended self is the principle that wherever a being's mental states are located or realized, there that being itself is located. He is essentially claiming that a corollary of the extended self is that a being's mental states must be located entirely within it (or at least if the extended self follows from extended mind) (485).

More of Olson's account will be explained later when an analysis of his arguments is provided. But I would just like to note the following. Before Olson concludes that the extended self implies that, "any psychological being could have parts located beyond its skin", he investigates three possible extended-self theses. Remember that one candidate was that, "it is possible for a human person to extend beyond her skin". Recall that, by discussing and refuting the broadest thesis he can thereby refute the narrow ones. While Olson does not explicitly state at any point that he is discussing the tenability of the extension of personal identity or persons, his admission that the rejection of one thesis will refute the others implies that questions about the extension of human persons are answered by questions about the extension of

psychological entities. Thus, even if he does not explicitly speak of persons here, he is nonetheless offering an analysis that rejects the extension of persons.

As we will see, Olson's article is an excellent addition to the literature that attempts to make sense of EST. Olson tries to make sense of exactly what Clark and Chalmers are positing in claiming that the self can be seen as extended. This being said, Olson still falls back upon the language of persons in making sense of their thesis. As will be addressed later, Olson talks about how EST can lead us to the conclusion that there is a difference between two entities: the initial organism and the subject of extension: there is, "a person who is not an organism and an organism that is (presumably) not a person" (487).

Olson is a reputable theorist in the literature surrounding personal identity and personhood, advocating animalism. Animalism is an analysis of personal identity that asserts that all humans, and subsequently all person stages of such, are biological organisms. While there is the potential that there may be non-human persons, Olson ignores these suppositions and focusses on what we as humans are: biological animals. Personhood, as he conceives it, is not a substance category (persisting in virtue of being a person) but a subsequent stage that is experienced by the human animal in its development. There are times when it seems that animalism is motivating his grievances with EST (See Olson 2011, 486). I will get to these grievances later on, but for now let us acknowledge how one of the most prominent authors who analyses EST does so within a framework that, at least minimally, presupposes that the notions of personhood and personal identity are informative in this investigation.

#### 4.2 Contextualization: Baker

Baker takes a more modest approach, instead positing that we can allow a modest thesis about extension while still maintaining that EST is false. Ultimately Baker posits that Clark and Chalmers set out a false dichotomy, in that we can reject both the implications of ‘shrinking the self’ and ‘seeing ourselves spread into the world’. Baker distinguishes between personal and subpersonal processes, arguing that subpersonal level processes can be considered extended. Baker distinguishes between the two by asserting that the personal level focuses on intentional agents and, “what they perceive and intend, believe and desire” (648). The subpersonal level is the level that which supports the actualization of personal level states. This level describes the vehicles and processes of cognition.

The distinction between the subpersonal and personal allows for Baker to posit a modest extended mind; while at the subpersonal level there may be the extension of cognition, personal level states are realized by persons and persons themselves are not the sort of things that extend. This allows for the logical move where she may claim that, “enduring persons may be subjects of mental processes constituted by extended cognitive processes” (650), while still claiming that persons themselves do not extend. Baker is going to hold to the claims that persons are not essentially biological, and that there are no immaterial minds. Specifically, the only coherent way we can understand ourselves is as entities with first-person perspectives.

Utilizing a framework she develops elsewhere for the constitution view of persons she is going to use this framework in analyzing and objecting to EST. Generally, the constitution view of persons that Baker articulates can be seen as saying that, “persons are material beings, wholly constituted by bodies—typically human organisms—but not identical to the bodies that constituted them. The relationship between a person and the organism whose brain makes possible the person’s thoughts is constitution, not identity” (650). The specifics of Baker’s analysis of personhood will be clarified later. For now, it is important to realize that Baker makes

sense of the shortcomings of EST through analyzing the concept of personhood against it. “The reason that I think that persons (or ‘selves’) do not fluctuate with various couplings is that the integration of parts of human organisms and machines takes place at subpersonal levels. At subpersonal levels, mechanisms and functions are explained in computational or neurophysiological or physical terms. At the personal level, we are focusing on intentional agents [and what they believe, desire, perceive and intend]” (649).

#### 4.3. Contextualization: Wilson and Lenart and Heersmink, Oh My!

Wilson and Lenart take a different approach. The two agree with Baker that persons are not the sort of thing that can extend by rejecting mental internalism; they claim that persons can have narrower boundaries than their mental states and cognitive systems of which they are a part. However, the two are going to allow for extended mental states and cognitive systems. Not only this, but they are going to advance a more radical claim that personal identity itself is best conceived of as extended: what makes a person the person they are across time is not just an internal criterion, but rather environment and externals help ground identity also.

Wilson and Lenart investigate memory specifically, given that memory plays a prominent role in discussions of both the extended mind and personal identity. Specifically, they focus on narrative or autobiographical memory, as they see this as the most relevant to discussions of personal identity. Wilson and Lenart claim, given the extended mind thesis, that, “narrative memory need not be bound exclusively to individualistic recollection but can come to incorporate the world beyond the individual in a variety of ways” (430). Essentially, what Wilson and Lenart are doing here is showing that memory is the sort of process that we can see in various ways as extended. Through a Lockean analysis of personal identity, which relies on the psychological continuity of memory, we can thus conceive of personal identity itself as

extended, in that some of the conditions that are required for this relation to hold are satisfied by extended mental states and processes. From now on, when I speak of personal identity as being extended I will have in mind the claim that the conditions required for it are satisfied by entities that extend beyond the skin-skull-boundary.

Regardless of the fact that personal identity can be conceived as being extended, they remain firm on the fact that persons, or individuals, are not the sort of thing subject to extension. "The characterization of wide realizations preserves the idea that properties with such realizations are still properties of individual subjects" (Wilson 2004); the externalist view of personal identity does not entail that the individuals who are persons are themselves wide or extended. The reason why they defend such a claim will be explained later. For now, let us just acknowledge more authors who make sense of the extended mind through personal identity and personhood, insofar as the authors acknowledge the impossibility of EST through the lens of extended persons.

Heersmink (2016) explores the implications of extended and distributed cognitive theory for our notions of personal identity. Ultimately he is going to argue that personal identity can neither be reduced to psychological structures instantiated by the brain nor by biological structures instantiated by the organism, but should be seen as an environmentally-distributed and relationally constructed (3135). He believes that the complex web of cognitive relations we maintain and develop with other externals (other people and technological artifacts) partly determines our 'self'.

Heersmink's contribution to the debates surrounding EST won't be discussed in significant detail in this chapter. Rather his work on the moral status of cognitive artifacts and the implications of extended and distributed cognitive theories on personal identity will be

discussed at length in the next chapter. For the time being, however, it is noteworthy that he takes the implications of extension to impact personal identity, articulating personal identity in a similar manner as the other authors discussed. “In this paper, I am assuming that psychological or cognitive properties are essential for both synchronic and diachronic aspects of personhood” (3141); “if our cognitive capacities are important for our personhood and if our cognitive [external] artifacts are ‘the things that make us smart’, then such artifacts are important for a synchronic conception of personhood” (3142). Heersmink understands the question of what defines a person as a question concerned with synchronic aspects. In this way we can understand Heersmink as trying to understand how the synchronic and diachronic questions can be made sense of through the framework of extension.

The above authors are not the only people to address the theme of extension in conjunction with personal identity and personhood but for the sake of space we will not endeavor to address the views of any more authors. If there is a significant takeaway from this section, it is not that EST can *only* be read through the lens of personal identity and personhood. But rather, the methodological framework surrounding personal identity and personhood is one such that it is well equipped to answer and tackle with the question of extension. As such, I believe that it is an apt endeavor to read EST through the lens of personhood and personal identity, and will continue to do so moving forward.

## 5.0 Direct challenges to EST

Regardless of a potential misreading of Clark and Chalmers, I am going to posit that a reading of EST through personal identity and personhood is still apt. Specifically, Clark and

Chalmers are concerned with how we ought to conceive of ourselves in the light of the extended mind thesis. Even though they attribute these questions of what we are to selves, the literature of personhood and personal identity was established to provide answers to these types of questions. It is the reason why so many of these authors have endeavored to answer the questions associated with EST through personhood and personal identity: because there already exists a vibrant and rich literature to pull from in explaining our intuitions. Thus, this thesis will go forth with the realization that there can be some complications in thinking that EST can be interpreted as a thesis about persons. This being said, the thesis will still endeavor to investigate the relationship between both personal identity and personhood and extension. What follows next is a critical analysis of authors who explicitly reject EST. The arguments presented by these authors will be investigated without the assumption that that they are arguing for the extension of persons unless they otherwise state that that is their intention.

### 5.1 Olson

Olson argues that the extended self thesis does not logically follow from the extended mind. Olson reads Clark and Chalmers as positing that, “the extended mind says that mental states and processes can be ‘extended’ or ‘external’ in the way that some of Otto’s memories are supposed to be” (483). From here he goes on to articulate two versions of the extended mind thesis. By virtue of the fact that Otto has to ‘look up’ the memories as a source beyond his own skin, this form of access gives the memories both a different *phenomenology* from ordinary memories and different *causal relations*, by virtue of their spatial configuration, to the rest of his mental goings-on. Of the two different extended mind theses, Olson focuses on the spatially extended mind in so far as it is relevant to the extended self and ignores the phenomenological version. Thus, we are lead to think of external mental states as those at least partly located or realized beyond their subject’s skin. This is not to say that the phenomenology of the extended

mind is not important here, but rather that the implications of the spatial configuration are all that matter with implications about EST. As Olson writes, "... there are at least two different extended mind theses, one concerning spatially external mental states and the other concerning causally and phenomenologically external ones. Seeing as only the spatial extended mind appears to be relevant to the extended self, however, we can ignore the causal and phenomenological version" (Olson 483).

Recall that Olson on to articulates the three different candidates for the extended self, focussing on investigating the possibility of the third claim, as proposing a legitimate rejection of the third implies a rejection of the other two. He defines a psychological being as, "a subject of mental states or a bearer of mental properties" (483), essentially entailing that any type of entity that can be considered to have mental states also has the capacity to possibly extend these mental states into their environment.

From here Olson goes on to note that Otto's description by Clark and Chalmers as 'a coupling of biological organism and external resources' suggests that he would have both the organism and the notebook as parts: he would be a disconnected or 'scattered' object. (484). Olson takes this principle to posit thus that any psychological being could have parts located beyond its skin. If Otto extends beyond his skin because some of his beliefs are located in his notebook, then the notebook is a part of him (484). This leads him to posit an implicit arbitrariness about how much of the notebook is a part of Otto: "which parts of the notebook would be parts of Otto according to the extended self depends on how much of the notebook his beliefs take up" (484). Olson posits that this question must have an answer given the theses of the extended mind and self.

It is unclear that this conclusion about the location of a border is acceptable. As the notebook acts as a repository for beliefs, and beliefs are, at least conceived of, as mental states constituted by physical processes but not necessarily reducible to them, inferring that beliefs take up a quantifiable amount of space seems vacuous. Thus, it seems that Olson needs to be committed to the belief that if we can ask the question and know how much of the notebook is included in Otto (since there must be an empirical answer), then similarly it is possible to physically quantify and isolate beliefs and other mental states. It seems like the interconnectedness of neurological and cognitive processes might not allow such a physical type of answer, in that multiple objects can be identified as activating cognitive processes that corresponds with belief formation. If we allow that belief states are individuated by content, not by spatial location, then we needn't reject EST since we cannot find an answer to the question of the physical location of mental states. The larger point here is a negligible one, but regardless I think that Olson claiming that we can empirically quantify how much of the notebook is part of Otto's extended mental states and mind is a *false* claim, and one that claims more about the world than the initial functionalist argument motivating the extended mind would infer. While Olson claims that the question of how much of the notebook would be a part of Otto must have an answer given by either the extended self or extended mind, I do not see why we must believe this.

In conjunction with the general principle that a thing is located where its parts are at, he posits that we might infer that Otto extends beyond his skin because a psychological being must extend as far as its mental states extend. This is the case because, if the extended mind was true but not the extended self, then it would be possible for a being's mental states to extend beyond its own boundaries, allowing the possibility for a mental state to be realized at least partly in things that are not parts of the being in the state. Thus, the link drawn between the

two theses is that a being's mental states are located or realized wherever the being is located. Mental-state internalism about the self, the position that a subject's mental states must be located entirely within it, follows accordingly. If the extended self follows from the extended mind thesis, then externalism about the subjects of mental states entails internalism about mental states themselves. In so far as one conceives that the mind can extend beyond the skin, then wherever these external mental states are, the mind must also be.

The positing of mental-state internalism is a critical premise in Olson's larger argument. He reads the extended self as implying that externals were not always a part of the extended subject; Otto's notebook was not always a part of him. In this sense, Otto was once confined to his skin; upon using the notebook, he gained new parts (485). Given Olson's reading of the extended-self as implying the accumulation of parts that are similarly viewed as containing the self implies, for Olson, that Otto is never a biological organism. "If (owing to the truth of the extended-self thesis) Otto's possession of external mental states implies that he is never an organism, then no one who has external mental states at some time is ever an organism. Having external mental states at some time and being an organism are incompatible properties" (486). Bear in mind, Olson advances an animalist view. So even though Olson goes further to show how unpalatable the extended self thesis actually is, this conclusion is already unacceptable from the animalist perspective. For the animalist, what we are first and foremost is human animals. To posit that we are something other than human organisms is to suggest we are not the thing that makes our personhood possible. By the animalist account, we are biological animals first, and the reason we are persons is because the biological animal has attained subsequent development to be characterized as having a person-stage. If the only thing we could ever be is a biological animal, then it seems impossible to reconcile the idea that the self,

the person-stage of the biological animal, is extended beyond the biological body (and for that manner, the very realistic boundary of where the organism and the environment are separated).

Here is the reconstruction of the initial argument posed by Olson:

P1. The EM thesis is true (external mental states are a conceptual possibility).

P2. Following from P1., external mental states are those at least partly located or realized beyond the subject's skin (s).

P3. If s, then any psychological being (Otto) could extend beyond its skin.

C1. Therefore any psychological being could have parts beyond its skin

P4. Otto's mind extends as far as his mental states extend

P5. Otto=Otto's mind

C2. Otto extends as far as his mental states (m)

P6. General principle: a thing is located where its parts are (p)

P7. Mental state internalism: a being's mental states must be located entirely within it

(q)

P8. If p & q and m, then Otto is not an organism

C3. Otto is not an organism

Roughly the argument proceeds from C3. into claims about the absurdity of this conclusion.

Before analyzing these conclusions, it will help to investigate some of the earlier claims Olson

makes in his argument. Earlier he claimed a general principle that a thing is located where and only where its parts are insofar as nothing can be literally located in a place where it has no parts (484). This is done to support the claim that wherever Otto's mental states are then he must also be there. This claim does a lot of legwork for Olson's argument. This claim is sensible when dealing with concrete particulars, like understanding the location of wooden tables. However, Olson has not provided an adequate framework of the persistence conditions of persons for us to understand how we may locate parts of persons. There is always the possibility that while persons supervene upon concrete particulars, they are ultimately not reducible to them. What is needed in future discussions of the extended-self thesis are rigid definitional frameworks that allow us to definitively say whether the self is something that can even be located within temporal space. It is unclear that supervening upon a human organism is enough to say that a human organism is where the self can be located. While this is not a failure upon Olson's part it is important to note how this point will influence future discussions of the extended-self thesis.

Another avenue for push-back is in complicating his claim for mental-state internalism. Mental-state internalism is the claim that one's mental states reside entirely within the organism. A complication comes not in accepting mental state internalism, but in questioning the motivation for the acceptance of this claim itself. Mental states, by Olson's account, are important to the psychological being insofar as they are parts themselves. Mental states are at least contained within the entity in that parts of the entity gave rise to the capacity to have mental states; if Otto is Otto's mind then Olson believes that Otto extends as far as his mental states do. It is here that we can run into some difficulty, for it is not clear that the mind is equivalent with its mental states. Is it the case that the mind is its mental states, or is it that mental states are constituted by the mind but neither reducible to the mind or definitive of the

mind itself? Is it possible to conceive of mental states as the product of the mind but not necessarily features of the mind itself? This discussion should be viewed with some skepticism, as very few terms here are substantively defined. Regardless of these difficulties, it remains an issue that will need to be solved if we are to be led to accept that external mental states can be legitimately viewed as a part of the cognizer.

There is another difficulty with mental state internalism. The question is whether we absolutely have to accept it. Wilson and Lenart provide reasons to think that mental states can extend beyond the individual, while still maintaining that it is the individual that is the subject of these mental states. Thus while the mind may extend, the concrete individual is not the sort of thing that does. Olson may push back against this by saying that what he is talking about here is psychological beings. Since mental state internalism can apply to psychological beings and minds, and individuals are psychological beings and minds, then thus we are led regardless into the absurd conclusions that the mental state internalism applies to selves. However, there is reason to at least question whether an identification can be made safely between psychological beings and selves or persons. A framework needs to be provided to prove that psychological beings are either constitutive of or identical to persons/selves, otherwise it may be the case that Olson's analysis does not go as far as he wanted. Recall that he proposed three possible readings of the extended self, going with the 'strongest' reading (that there exists a psychological being who could extend). If it turns out that we can dissociate psychological being from persons and selves, then it is not immediately clear that the success of his argument guarantees the impossibility of EST or extended persons. Thus even if mental state internalism leads us to the conclusion that psychological entities extend as far as their mental states extend, it is not clear that the identity relation is one such that a person or self extends as far as the psychological being does.

The conclusion that Otto is not an organism raises more than just one question. We are forced to posit Otto as a malleable entity with no fixed form as well as positing his body, O, a thing that is necessarily capable of thought if it is possible for Otto to 'recede' into his skin. This seems to also raise the problem of too many thinkers, where there is a thinking agent Otto and there is O that is both capable of thought. "But if I am not an organism and my animal body is something other than me, and that organism thinks just as I do, then I am one of two beings thinking these thoughts and writing these words" (487). I am not so sure that this response is as problematic as it need be if we grant the non-reductivist approach proposed earlier. If we admit that I am the organism but that the organism is simply a part of me and I am not reducible to the organism, then the fact that the organism is capable of thought does not commit us to dissonance since I am capable of thought. An interesting way that Olson might reply to this claim is with the following characterization. "The extended self implies that O, the organism associated with Otto, is not a subject of the external mental states located in Otto's notebook. They are Otto's beliefs but not O's. So although Otto believes that the museum is in Bloomsbury, O does not: at most O may believe that the museum is wherever the notebooks says it is" (487).

I think one way to respond to this claim is by appeal to personal and subpersonal systems, as articulated by Daniel Dennett in explaining pain. Subpersonal levels of explanation involve a scientific account of the various neurophysiological activities triggered by afferent nerves responding to damage. The subpersonal level of explanation does not need to actually reference the phenomenon of pain, but rather explains the physical behavior and information processing that the system responsible for registering bodily damage carries out. The personal level of explanation explicitly uses the concept 'pain' and applies it to the individual as a whole. I think that it is possible to apply this level of analysis to the systems that are involved in the

instantiation of Otto's belief. Just as the activation of the visual cortex is not perception as we characterize it, O is not enough alone to explain the access to the occurrent belief Otto has in his notebook. Rather it is the self, Otto as an entity beyond his body, that realizes the belief. In this way we don't need to expect that O shares Otto's internal mental states entirely, just as the notebook cannot be expected to share in all of Otto's internal mental states even though it is considered to be part of Otto, and just as the activation of the visual cortex is not all there is to the instantiation of perception.

Olson's argument raises some interesting points for future considerations of extended personal identity. Humans typically experience their subjectivity from the point of view of their body. Even if the subjective quality of their experience is not pertaining to their body, the argument can be made that the causal machinery necessary for the experience of their subjectivity is only possible within the brain. But with this said, is it necessary that the body in its entirety be considered part of one's extended personal identity? The borders can become vacuous when people consider bodily processes that are automatic and uncontrollable (such as feeling pangs of hunger or experiencing unmotivated desires that were created through associative structures). Even more interestingly is how, even though the focal point of experience might be limited to one's body one is able to embody phenomenological states correlated with one's social environment. One may feel anxious and distressed from considering the position of another person, or feel responsible for actions in which they have no sense of agency for. There seems room to question whether the human organism is the only thing we base our identity off of, considering that there are aspects of the body that we do not want to consider in our analysis of personal identity, and that likewise there are aspects external to the body we may want to include.

Of course, Olson may respond by saying that it is all well and good that we want to articulate our identity away from the organism, but that ultimately what we are is human animals, and that the longing for personal identity is an aspect of the human body actualizing a person-stage. As it stands, there should be good reason to think that one may still advance the possibility of the extended-self thesis (or at least the thesis for extended-personal identity). However, there is good reason to be skeptical of the reasons provided for refuting Olson. For one, this thesis has not posited an analysis of personal identity or personhood on which to base my characterizations of personal identity as a concept. Thus, when claims are made about the non-reducibility of personal identity or personhood to parts, the claims come contingent upon a conception of personal identity that does not necessarily align with specific analyses. These considerations will be strengthened with future discussions of personal identity in the fifth chapter.

## 5.2 Baker

Baker's framework differs from Olson's in important ways, and is ultimately motivated by a constitution view of personhood, a view that allows for different conclusions when addressing the EM thesis. Baker views thoughts and other contentful states as having two kinds of properties: properties determined by the content and properties of the vehicles that carry content. This differentiation has two interesting implications. The first is that it is by virtue of the content that we are able to individuate between contentful mental states. The vehicles of these contentful states do not play any role in individuation. The second point is that there is no systematic relation between the vehicle and the content of our thoughts. The example that Baker gives is of a highway sign. There are certain constraints on the kind of things that can be highway signs, but within these broad constraints many things may be a sign. In the same way there is the possibility that different vehicles can constitute the content of thoughts, and that

these vehicles largely do not play a causal role with respect to the effects of our thoughts. Like Olson, she differentiates between internalism and externalism of mental states. Baker noticeably differentiates between internalism and externalism of content and of vehicles of contentful states. Thus she sees the EM thesis of Clark and Chalmers as an extension of the externalism of contents to an externalism about vehicles.

Baker believes that Clark and Chalmers largely set out a false dichotomy in their characterization of extended selves: it is not a matter of having to choose between shrinking the self to a bundle of occurrent states or seeing ourselves as 'spread into the world'. Thus she raises two qualms with Clark and Chalmers concerning the nature of human persons. The first amounts to the same form of objection Olson puts forward. She posits that while there can be cognitive systems that have as parts Otto and his notebook, "Otto does not expand to become an extended entity that includes his notebook. Otto the human being does not dissolve or disappear into a cognitive system" (647). These claims are motivated by the idea that Otto is a concrete particular, but cognitive systems are not concrete particulars. A person who is part of a cognitive system does not expand to include the other parts of the system; nor does the system take ontological precedence over Otto the human being.

The second qualm Baker raises concerns the distinction between personal and subpersonal levels, a distinction that Baker takes to be ontological. She believes that the extended mind blurs this distinction, in that there is some ambiguity in regards to the things that extend cognition. For example, a neural implant is at the subpersonal level, but a tool is less clearly at the personal or subpersonal level. In response to this qualm Baker proposes a modest hypothesis that, "recognizes us as enduring persons whose subpersonal states may have nonbiological parts that play essential roles in cognitive processing" (648). In making sense of subpersonal and personal level states, Baker makes the following distinction: mental states and

mental processes refer to person-level states and processes that are constituted by subpersonal states and processes. Cognitive states and cognitive processes refer to subpersonal states and processes.

There are a few other important details Baker makes about the distinction between personal and subpersonal states, as it is these claims that will ultimately motivate her modest hypothesis. Baker claims that the subpersonal vehicles of a persons (person level) mental states are causally integrated with the person's other subpersonal parts and that some people have mental states that have subpersonal vehicles with nonbiological parts. Subpersonal states are cognitive processes and can be instantiated in multiple ways. At the personal level, we are focusing on intentional agents and, "what they perceive and intend, believe and desire" (648). Baker posits that a state is at the personal level if the person can come to acknowledge the state as her own (649), and that consciousness is not always a reliable measure of whether a state is at the personal level. For instance, an unconscious desire that a person can bring to consciousness as her own is at the personal level. Personal level states are concrete particulars, whose identity is determined by their content. This content is not determined by the vehicle, and these vehicles are not necessarily subsumed within the boundary for the person and their bodies (652). Importantly, Baker is making the claim that conscious acknowledgement of a state is sufficient for the state to be at the personal level, and also that conscious acknowledgement is necessary for the state to be at the personal level; "any state that does not presuppose consciousness is at the subpersonal level" (649).

Her argument for the modest hypothesis is advanced with her analysis of persons. Baker advances a constitution view of persons: persons are material beings, wholly constituted by bodies—typically human organisms—but not identical to the bodies that constitute them (650). For Baker, the relation between a person and the organism whose brain makes possible the

person's thought is constitution, not identity. Persons naturalistically arise through certain measures of constitution: when a thing or property of one primary kind is in certain circumstances, a distinct thing or property comes into existence or is exemplified (650). The person and the organism thus have different persistence conditions, thus the organism is essentially biological but the person is not. It is not that the person is not ever an organism, but rather that it is possible that the subpersonal processes underlying the person can be conceivably bionic. A person can have subpersonal parts that are not organic, and these can be material vehicles (or components of vehicles) of a person's mental states. In this way, the person can have extended cognition, because the material vehicles of her contentful mental states may be inorganic. But the person is constituted by a body, perhaps partly bionic, and the person spatially coincides with the body that constitutes her and does not extend beyond it. In this way extended cognition, the utilization of external subpersonal processes, is possible, but persons themselves cannot extend.

One initial worry for Baker's view is that it is not a definitive rejection of the possibility of extended-persons. She states that persons are constituted by bodies and spatially coincide with that body. If it is possible that the body can be altered to be inorganic, then it is possible that one might be constituted by an entirely inorganic body. The persistence conditions of an organism entail that the detachment or extension of certain parts of the body may be incompatible with persistence. But it does not seem that an entirely inorganic body might not have persistence conditions that are incompatible with extension. If we consider technologies like cloud storage and allow that an inorganic body might be able to make usage of this kind of inorganic storage, then it could be argued that persons cannot currently extend but could conceivably. Baker herself talks about how subpersonal states and processes may have bionic components. "On my modest hypothesis, subpersonal states and processes may have bionic

components... I take minds to be at the personal level, constituted by brains or by brains-with-bionic-parts (or someday, perhaps, wholly by bionic mechanisms)” (Baker 648).

A few other problems present themselves when we consider Baker’s characterization of personal level states. Baker says that a state is at the personal level if the person can come to acknowledge the state as her own. Not only does it seem that acknowledgement can be used as a tool to identify certain states at the personal level, but more strongly that the mere act of acknowledgement is sufficient to make the state a personal level one. The reading of this type of recognition of states is not inaccurate, as Baker says later that, “an unconscious desire that a person can bring to consciousness as her own is at the personal level. Any state that does not presuppose consciousness is at a subpersonal level” (649). This claim that the acknowledgement of states by the individual makes them personal level can be problematic, as it is possible for individuals to acknowledge mental states as their own even if they are not actually in them.

There are reasons that one may be critical of the claim that acknowledgement of a state is enough to bring it to the personal level. Confabulation, as understood in the scientific literature of memory impairment, is the production of statements and actions that are unintentionally incongruous to the patient’s history, background, present and future situation. This disorder has been a sign of Korsakoff’s syndrome (Benson et al. 1996), but has also been observed in subjects suffering from a myriad of other neuropsychological conditions (Barba et al. 2017), as well as, on occasion, in subjects who are neurophysiologically typical (Barba et al. 2017; Colheart 2017).

Within the literature of confabulation, a distinction is made between two types of confabulation Spontaneous confabulation reflects the production of an “incoherent and context-free retrieval of memories and associations resulting from the superimposition of

frontal dysfunction of an organic amnesia” (Barba et al. 2017). Spontaneous confabulation is confabulation that is offered by the patient. Provoked confabulation is confabulation not spontaneously produced by the amnesic patient but instead produced only when the patient is asked to perform some task (Colheart 2017) and largely reflects a normal response to a faulty memory. An example of provoked confabulation is of a subject being told a brief story then retelling the story later with new or altered material in the story recall. An example of spontaneous confabulation is given by Kopelman (1967) who refers to one patient who talked incessantly about her mother (who had in reality been dead for many years). The patient would explain to the interviewer that she had to hurry to cook dinner for her mother, and that she had been ticked off for being late in preparing the dinner the day before. Upon being asked whether her mother was still alive, the patient responded forcefully that her mother was very much alive. The patient was confabulating not only that she had desires to cook dinner for her mother, but was also confabulating non-existent emotional states.

The main point behind this example is this: the subject who would exhibit signs of conflation about feeling angry at being late to preparing dinner for her mom is clearly acknowledging a state that she is not actually in. Thus while she is acknowledging the state as her own, this does not seem enough to bring this state to the personal level. This shows that, with patients that are falsely attributing states to themselves, that the acknowledgement condition does not seem to be enough to bring a state to the personal level. Thus the sufficiency claim that Baker makes, the claim that a state is at the personal level if the person can come to acknowledge the state as her own, does not seem to be a valid claim that will capture the recognition of a state as personal or subpersonal.

The point of this discussion is that Baker criticizes Clark and Chalmers for creating some ambiguity in the distinction between subpersonal and personal level states. Yet Baker’s account

of these states faces its own challenges and, by certain readings, can actually lead into a misconstrual of what actually differentiates between these two types of states. The tenet of Baker's modest proposal is that subpersonal states can be constituted by non-biological parts. If we cannot conceive of a more systematic way to differentiate between these two other than by conscious acknowledgement then one might be led to a confusion about what actually extends.

Baker argues that what makes the thought or contentful state the very one it is its content, not what constitutes it (not the vehicles). The content of a thought may depend on all manner of vehicles that constitute it, and it may even depend upon features outside 'the skin'. Certain thoughts cannot exist in the absence of complex relations to the environment. This kind of claim does not defeat Baker's modest proposal. The content may be contingent upon externals, but the content is still constituted in an appropriate way that we can think of these personal level states as not extending. But this claim can open up the path towards an alternative conceptualization of the extended-self thesis: a weaker claim that posits that certain necessary processes and states that constitute the person themselves are external to the body. Persons may not metaphysically extend in the way that Clark has in mind when he says that, "the self is a hastily cobbled together coalition of biological and non-biological elements, whose membership shifts and alters over time and between contexts" (2004 177). Even if persons are not the sort of things that logically can extend, it is still perhaps possible to see identity as being importantly constituted by one's externals. However, if we are inclined to push back against Baker's arguments here then I believe we have good reason to accept a stronger claim, the claim that the tenability of the extension of persons and of personal identity is still possible even in the face of these objections.

## 6. Conclusion

With the above discussion, it seems clear that we have reason to question Olson and Baker in their conclusions that persons cannot extend. This does not amount to the conclusion that persons or personal identity do extend, but rather that future discussions into the possibility of extended persons and personal identity can be undertaken with legitimacy. In summarizing our analysis of their arguments, I would like to note the following. The discussed authors present direct arguments against EST. For the sake of space only these two authors were addressed, albeit there are other others who indirectly object to EST. One notable author is Tom Buller (2013). Buller's criticism is that conscious experience is necessary for personhood and cannot be extended, thus personhood should not be conceived of as extended. Most accounts of personhood would be remiss if they did not make sense of how conscious experience fit into their analysis of persons. Generally, the view is held that most human animals are persons. Likewise, many non-human animals are not considered to be persons. If we could infer that these non-human animals are incapable of conscious experience and that human animals are, then perhaps it is conscious experience that aids in the status of personhood? This type of claim is the sort of burden that accounts of personhood typically will respond to, or at least provide some semblance of an answer to. Without getting to bogged down in discussions of what entails personhood or personal identity, there is reason to think that conscious experience is not all there is to an adequate articulation of personhood. While conscious states may be relevant for personhood, it could be claimed that we are missing some other relevant and necessary aspects of personhood by ignoring, say, non-occurrent beliefs or non-occurrent information in memory (Heersmink 3141).

Many analyses of personal identity and personhood deal with problem cases and thought experiments in order to advance our understanding of these concepts. For instance, Parfit talks of fictive examples of malfunctioning teleportation and brain fission in order to

advance our knowledge of how we can conceive of the relationship between survival and identity; Locke talks of body swapping in order to understand the importance of psychological continuity in supporting our intuitions. Recall that one of my conclusions in the first section was that, in regards to how we understand the cognitive, it may be that we are held to stricter regulations in what we deem cognitive. While Martian intuitions may be informative, it may be that we mischaracterize what is cognitive in some instances. The literature on personal identity and personhood, however, does not face such regulations. In fact, in many instances it is the expansion of our existing definitions of the two concepts through special cases that allows us a better and richer understanding of the nature of personal identity and personhood.

Here comes an important difficulty for this thesis. Recall that HEC ought to be grounded in some 'realism' about the cognitive. Likewise, the extended mind and EST remain contingent upon HEC in important ways. Thus an important difficulty arises: in conceiving of the feasibility of EST, should we utilize thought experiments and abstract thinking about the 'possibilities' of selves or persons? Or are we restrained to a degree by how selves are conceived of within a contemporary context (the idea here is that we can think of selves as they currently exist, without the aid of thinking about possible futures or experiments that would alter the self)? Or, perhaps more specifically, are we at least constrained in the way we can conceive of cognition in these thought experiments? These are important questions for this thesis, for it may turn out that some of the conclusions I come to later in the thesis are untenable because of an unrealistic conception of cognition. Thus while we may have a valid conclusion about the extension of persons or personal identity, it may turn out that EST is falsely shown as viable. This possible conclusion is granted if we allow that EST cannot be verified with special cases. Yet we may not necessarily be forced into this position if we allow that EST is best, and perhaps only understood

through the framework personal identity and personhood. The only constraint that may apply is that we need to be careful with how we characterize cognition.

This leads into a larger question pertaining to extended personhood and personal identity, as will be investigated in coming chapters. Is it that the necessary and sufficient criterion of personhood and personal identity must be realized within the same system, or even at the same time, or can we possibly notice the realization of these different criterion under varied conditions? Are these criterion to disparate at this point if they are not realized under similar conditions? Or is it that these criterion must be realized coextensively in order to even be considered criterion in the first place? There is more work to be done to establish what exactly can legitimately be entailed when we consider the extension of personal identity and personhood, as will be done in the coming chapters. For the time being, however, let us at least conclude with the knowledge that the objections against EST are not as damning as we may think, and that we may proceed forward in our investigation of extension without undue concern that these themes are improbable from the offset.

## 1. Introduction

The last chapter concerned itself with analyzing various authors who reject the possibility of EST. After brief exposition and investigation, the claim was made that the authors who reject the possibility of EST do so hastily: their arguments are not enough to definitively state that EST is implausible from the outset, and likewise that there is significant room to push back against these and other authors who reject EST. The conclusion left with the reader was that EST, as it stands, is not implausible as these authors argue, and thus that investigations should be undertaken to analyze the possible reception of extension.

Prior to this was an analysis of the original articulation of EST by Clark and Chalmers, and a subsequent discussion of how we ought to proceed in our analysis of EST as they articulate it. Following a brief exploration into the concepts of personal identity and personhood, it was claimed that analyzing EST through the lens of personal identity and personhood was an apt and advisable endeavor, considering that the literature surrounding the two concepts is detailed with the contributions of many philosophers who attempt to answer abstract questions about the conditions that entail personal identity and personhood. Likewise, the literature strives to conceptualize the nature of what we are and what ensures persistence, notions that EST provides implications for. As such, the thesis moved forward with the understanding that we would analyze questions of extension through the lens of personal identity and personhood, thus questioning whether we can consider one's personal identity or a person as extended beyond the skin-and-skull boundary.

This chapter begins what is the first substantive positive push in this thesis. Rather than analyzing and disagreeing with authors who reject extension in its various incarnations, the case will be made here that extension is tenable. This project will proceed by undertaking a

systematic investigation of the role of memory systems for personal identity. As it will be shown, various authors who subscribe to psychological criteria of personal identity and personhood make explicit reference to the role of memory in grounding the identity and status of a person. Through the clinical analyses of Addis and Tippet (2004), the argument that memory is essential to a sense of identity, granting a psychological approach, will be strengthened. Following the earlier claims that Otto uses his notebook as an external memory bank, and that his utilization is functionally similar to Inga's, then the claim follows that the extension of the type of memory that is relevant to personal identity and personhood is plausible. This conclusion will likewise be strengthened by considering authors who investigate the nature of autobiographical memory in relation to the question of personal identity and personhood, as well as by analyzing the role of different forms of non-autobiographical memory in constituting personal identity and personhood.

The contribution I want to make above a reading and demonstration of the extension of personal identity and personhood through the criterion of memory is, more generally speaking, that the consideration of the extension of personal identity and personhood is a feasible and apt endeavor. This chapter focuses on how psychological accounts can, through the role of memory, be considered as extended. The next chapter will concern itself with analyzing other contemporary accounts of personal identity and personhood within the context of extension, showing various ways that we can consider extension beyond the perspective of psychological and cognitive extension<sup>13</sup>.

---

<sup>13</sup> The idea is that, many authors defend the extension of personal identity and personhood through appeal to and immediate acceptance of Otto and his notebook. This thesis has sought to provide grounding for this appeal by analysing HEC and EST in rigorous detail, discussing objections and responding to them. What I want to do is argue that the very tenability of extension of the concepts of personal identity and personhood is possible, independent of the reception of a specific analysis of those concepts.

The concepts of personal identity and personhood are supported by analyses that argue for their accurate descriptions. The reason why one is compelled to accept a psychological analysis rather than a bodily analysis is through appeal to intuitions. If one agrees with the intuitions put forward via the prince and the pauper case, then one may support Locke's analysis. Likewise, if one agrees with the intuitions behind the brain transplant via a mad scientist case, then perhaps one will lean in the other direction towards a bodily criterion<sup>14</sup>. These disputes between competing analyses are done through actual and fictional examples that examine our concepts of personal identity. How well these theories can respond to these problem cases strengthens the view of the specific analysis.

If we can grant that personal identity and personhood are the sorts of things that can extend, wouldn't views that support this notion only be seen as stronger? Yes and no. If one follows the arguments presented in this thesis, then one may be of the opinion that extension is tenable and plausible. It may even be that we come to the conclusion that Otto's personal identity is extended into the notebook. If we think that the conceptual possibility is enough to think that this type of extension is possible, then we might think that analyses of personal identity and personhood that do not accommodate extension are thus not apt analyses of these concepts. However it is not necessarily the case that this theory sides with common sense intuition. Consider the fact that most analyses of personhood recognize humans as persons. While many analyses have alluded to the possibility of non-human persons, most proceed as though the only persons in existence are humans. As humans, we experience the world and

---

<sup>14</sup> These examples are used by Locke and Williams respectively to support competing analyses of personal identity. The Prince and the Pauper is essentially a thought experiment where two individuals experience the swapping of their psychological contents. Locke claims that our future concerns about continuity and identity would logically follow with the psychological continuity, thus supporting the psychological analysis of personal identity and personhood. Williams does something similar, instead describing the thought experiment of 'body-swapping' in a different way that emphasizes the body in our considerations of future concerns. Both of these views will be discussed in more depth in chapter 5.

ourselves from the same spatial body we were born with. Our existence is characterized by our embodied condition. It thus seems implausible to say that someone is something beyond their body, beyond the skin-and-skull boundary. The intuition by many thinkers is that extension of personhood and personal identity is implausible. It is probably even the case that the readers of this thesis considered this project a push against the existing methodologies and literature of personal identity and personhood. Would it be a strength of an analysis if it were compatible with extension, or even a weakness? These are all methodological questions that may not be definitively answered.

I would like to note one thing, however. While analyses of personal identity and personhood may be gauged on the possibility of extension, it may not be the case that these analyses must accommodate extension in order to ensure that they are viable analyses. For it is the case that some analyses simply cannot accommodate the possibility of extension. This may be because personal identity and personhood is conceived of through a bodily criterion, that it is only organisms that can realize persons, etc. We may find reasons to be critical of these analyses because of this, but we must also provide other argumentation to show why their conception of personal identity or personhood is not apt. Otherwise, we may end up mischaracterizing personal identity and persons. For instance, I could posit an analysis of personal identity right now that entails extension. Let us grant some analysis where one's property is conceived of as part of the person, perhaps adopting aspects of Locke's theory of property and allusions to the state of nature. Under such a conception, one comes to own property upon infusing the material world with their labor. If we allow this, we might posit an analysis of personal identity where one's person extends into all property they own. In evaluating this analysis, we may say that it has the strength of accommodating extension. However, this theory mischaracterizes what it is to be a unique individual. If we were to take away an apple that Shelton himself

picked, we are not limiting his mind or fundamentally impoverishing his identity. While this theory can illustrate extension, it does so in a way that mischaracterizes the concept it sought to explain. If one allows the work done in this thesis, then it is plausible and even probable that personal identity and personhood can extend, and that with a bit of legwork we can understand many analyses of personal identity and personhood as extended.

What follows is a short discussion of the role of memory in discussions of personal identity and personhood.

## 2. Preliminaries

One way authors have understood personal identity and personhood is through a psychological approach. These sorts of accounts gauge diachronic conditions through assessing whether certain mental features (things like beliefs, memories, capacity for rational thought, etc.) survive. Synchronic conditions are concerned with the necessary capacity and actualization of these mental features. These kinds of accounts would take the following form:  $x$  at  $t_1$  is  $y$  at  $t_2$  if and only if  $x$  is psychologically continuous with  $y$ . It is this sort of biconditional that motivates various analyses that fall under the larger umbrella term of psychological accounts.

Harry Frankfurt (1971), for instance, advocates a volitional view of personhood. Frankfurt believes it is the actualization and retention of second-order volitions that informs our diachronic and synchronic concerns. Frankfurt's view, in short, is that volitions are types of second-order desires. A second-order desire is a desire about a desire, taking the form of  $x$  wants to want to do  $F$  (Shelton wants to want to press a button). A second-order volition takes the form of a second order desire with an added feature: second-order volitions are committed to the first-order desires contingent upon themselves. In such cases the statement that "A wants to want to  $X$ " means that A wants the desire to  $X$  to be effective in moving themselves.

“He wants this desire to be effective—that is to provide the motive in what he actually does” (Frankfurt 10). Thus this entails that the volition about the first-order desire have the first-order desire actually be wanted. Second-order desires simply pertain to the wanting of a certain desire without presupposing that the agent wants the desire to be effective. Second-order volitions entail their subsequent first-order desires and that these desires act as a motivating force. Thus we can think of a second-order volition as being a desire that one’s desire be a motivating force in one’s actions, e.g., Fred wants his desire to study to be the desire that motivates him in his actions. It is the possession of second-order volitions that Frankfurt regards as being essential to being a person (10).

Since volitions are conceived of as desire states, and desires can be conceived of as mental states, we can conceive of his account of the synchronic conditions of identity as being concerned with mental features and their survival. While Frankfurt’s analysis has been conceived of as a narrative analysis<sup>15</sup>, for the sake of our investigation we simply want to note this analysis as focusing on mental features and states: we can only understand Frankfurt’s analysis through understanding the role of desires in his framework.

Other authors have used variant psychological views as a grounding for their conception of personal identity and personhood (Parfit 1984, Lewis 1976, Shoemaker 1970, Nozick 1981, Nagel 1986, etc). Specifically, some authors have used memory as a criterion for understanding persistence conditions of identity. For example, Locke’s account of what makes a person appeals to consciousness, and says that the persistence conditions of the identity of a person are guaranteed through a form of psychological continuity. He writes, “For since consciousness

---

<sup>15</sup> Narrative accounts of personal identity and personhood posit that individuals form identities by integrating their life experiences, beliefs, memories etc. into an internalized, evolving story of the self. Persons, on this account, are informed by this narrative as either the system that made this narrative possible or as the narrative itself.

always accompanies thinking, and it is that which makes everyone to be what he calls self, and thereby distinguishes himself from all other thinking things: in this alone consists personal identity, i.e., the sameness of a rational being; *and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it, that that action was done*" (Locke 39, emphasis added).

Locke uses this criterion of memory to make sense of his thought experiment, that of the prince and the pauper. Upon the reveal that the two have had their bodies switched, Locke answers the question of identity through pointing out that the conscious subjects still have access to the memories they had when they resided in their original body. "For it being the same consciousness that makes a man be himself to himself, personal identity depends on that only, whether it be annexed solely to one individual substance, or can be continued in a succession of several substances. For as far as any intelligent being can repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present action; so far it is the same personal self" (40). Thus it is for Locke that it is consciousness that grounds our ascription of personhood, and it is the ability of an individual to consciously access their memories that guarantees persistence.

Take another analysis that views psychological connectedness as the criterion that grounds identity. Derek Parfit (1984) posits the following analysis for his conception of personal identity:

The Psychological Criterion: (1) There is psychological continuity if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity

has the right kind of cause, and (4) there does not exist a different person who is also psychologically continuous with Y. (5) Personal Identity over time just consists in the holding of facts like (2) to (4). (Parfit 207)

Psychological continuity, for Parfit, is established by the inheritance of beliefs, memories, preferences and other mental features of x by a surviving entity y. As such, we can understand his view as holding to the importance of memories in grounding the psychological criterion, among other things. Likewise, Parfit (1971) notes in an earlier piece that, “the most important particular relation is that involved in memory” (14). In a footnote discussing philosophers that hold the belief that memory is important in relation to personal identity, Parfit notes that, “those philosophers who have held this belief, from Butler onward, are too numerous to cite” (14). As such, we will go forward with the knowledge that memory is one of the most popular criteria in establishing personal identity.

The main concern of the above was to illustrate the role that memory plays for many psychological analyses of personal identity and personhood. One could make the larger claim that, beyond specific analyses that explicitly posit memory as a criterion for understanding personal identity, memory often plays an implicit role in accounts of personal identity. For instance, if one subscribes to a volitional account (Frankfurtian) of personhood, one believes that it is the retention of a volition or the structure necessary for the articulation of volitions that entails personhood. However, it can be argued that the retention and continued existence of a volition relies upon an established memory system to: 1. Ensure that the volition is kept and updated by the individual in the face of their evolving life, and 2. Have autobiographical memories in place to provide a basis for the individual to assert a volition that makes sense according to his experiences and desires.

Some authors may push back against this claim for the significance of memory. Consider Olson's specific take on animalism (1997), the view that we are human animals first and foremost. For Olson, it does not matter whether the animal at t2 retains memories from t1 to ensure that it is the same animal at t1. Rather, it is in virtue of being the same organism that identity is ensured. Against these views the argument for the importance of the criterion of memory can be argued. However, there is only so far one can go in attempts to consolidate the analysis with the importance of memory before the analysis loses its original flavor. Against these analyses it may be the case that the importance of memory is realized but is not accommodated as a necessary or sufficient condition in the analysis. At the very least I would like to advance the claim that memory seems a good criterion for grounding personal identity, and that, when considering analyses that do not emphasize memory, we should consider whether the analyses can provide equally strong motivations as memory normally does.

In what follows I will describe the experimental paradigm of Addis and Tippet, who show through a clinical study the importance of autobiographical memory for identity. I will argue that this study supports the intuition advanced in this chapter that memory is important for diachronic and synchronic concerns, irrespective of specific analyses of personal identity and personhood. Prior to this I will provide a working clinical definition of memory to be used going forward.

### 3. Memory and Addis and Tippet

In conceiving of memory, I draw on the work of Klein and Nichols (2012) and upon the work of Addis and Tippet. In distinguishing between memory systems, Klein and Nichols note the distinction between procedural and declarative memory, as most psychologists agree that, at its most general level, long-term memory can be characterized as containing two basic

systems: procedural and declarative memory systems (Klein and Nichols 2012; Cohen 1984; Tulving 1983). Procedural memory makes possible the acquisition and retention of motor, perceptual and cognitive skills; it consists in the nonconscious expression of previously acquired behavioral skills and cognitive procedures. Declarative memory consists in facts and beliefs about the world (e.g., knowing that canaries are yellow; knowing that Shelton likes to push buttons). It is the two subsystems or types within declarative memory that are relevant to the investigations of the authors, these subsystems being episodic (or personal incident) memory and (personal) semantic memory<sup>16</sup>. These authors both look at the role autobiographical memory plays in relation to identity, and thus focus on these two subsets of declarative memory.

Episodic memory, as articulated by Klein and Nichols, records events as having been experienced at a particular point in space and time. Episodic memory represents the what, where, and when of an event; it is experienced as a memory that makes explicit reference to the time and place of its acquisition. This subtype of declarative memory has been thought to involve re-experiencing events from one's past, thus providing its owner with content by which he or she is able to construct a personal narrative, that is, his or her life stories (Klein and Nichols 679). As articulated by Addis and Tippet, personal incident memory refers to the episodic memory component of autobiographical memory, that is, memory of a specific personal event (57). Semantic memory refers to relatively generic, context-free information. Importantly, semantic memory typically lacks a source tag—it does not specify when or where

---

<sup>16</sup> Addis and Tippet diverge from Klein and Nichols only insofar as they define the two subtypes of declarative memory in such a way. For Addis and Tippet, it is both personal incident memory and personal semantic memory that they evaluate in their experiment in order to assess the disparity between participants with Alzheimer's disease and control participants of the same approximate age demographic. While they will focus on the articulation of autobiographical memory, they refer to both subsystems of declarative memory (personal incident memory and personal semantic memory) as playing important roles.

the memory was acquired; rather, it is retrieved as knowledge without regard to where and when that knowledge was obtained.

Addis and Tippet make the claim that, “autobiographical memory is thought to contribute to trait self-knowledge and to self-narratives, enabling the integration of past and present selves and contributing to the sense of continuity of identity” (56). With the view that autobiographical memory contributes to identity leads to the prediction by the authors that a loss of autobiographical memory results in changes to one’s sense of identity. Thus Addis and Tippet conduct a study to investigate the status of autobiographical memory and the sense of identity in individuals with Alzheimer’s disease (AD) and age-matched control subjects, while also investigating whether degree of autobiographical memory impairment is associated with changes in a number of aspects of identity. Addis and Tippet note that significant impairments of personal incident (episodic) memory have been documented in many major studies investigating the status of autobiographical memory in AD individuals. The current literature, as assessed by the authors, indicates that AD individuals are likely to have impairments of autobiographical memory, and therefore these patients may provide a suitable population with whom to investigate questions regarding the relation between autobiographical memory and identity.

The authors understand the term ‘identity’ interchangeably with self-concept. Identity for Addis and Tippet is understood as a subcomponent of self-concept along with self-esteem and behavior. Identity can also be understood as self-knowledge, which refers to self-identifications and perceptions about oneself. Likewise identity for Addis and Tippet can be understood through personal narratives, through the stories one constructs about oneself, one’s life and experiences. Primarily, Addis and Tippet believe that we can gauge the strength of identity through three criteria: multidimensionality, coherence and continuity over time. The

authors claim that identity is generally thought to be multidimensional, and thus believe that they can gauge the strength of identity of an individual by evaluating the different domains that aid in the consolidation of identity. These dimensions may include personal, physical, social, family, different identities for different social contexts, or different identities over one's lifespan. In other words, one's identity ought to be informed by many different contexts and domains, and thus the more dimensions that an individual can point to as being important for who they are strengthen the overall character and strength of one's identity.

Yet despite being multidimensional, identity has an overall coherence which gives a sense of unity rather than the subjective perception that one's identity is fragmented (Addis and Tippet 57). If coherence is the measure of how well unified these different domains are, then continuity over time is the measure of the perception that an individual is essentially the same person they were in the past, despite changes in their roles and life-situations. This measure of continuity is achieved through evaluation of the integration of past, present, and future selves. Addis and Tippet thus claim that autobiographical memory impacts upon these different dimensions in important ways, supporting the perception of proper multidimensionality, coherence and continuity over time. Thus they claim that autobiographical memory has a clear role in the constitution of identity, enabling connections between discrete moments in consciousness, and, if lost, it would follow that one's identity would also be lost or significantly changed (57).

Before moving forward, I would like to note the following. It appears, by the authors characterization of identity, that they are not explicitly talking about the philosophical concept of personal identity. And if this is the case, then while this experiment may prove that the loss of autobiographical memory would result in changes to one's sense of identity, then it is not necessarily that the loss of memory impacts upon personal identity. This sort of conclusion can

be drawn by analyzing the authors' characterization of multidimensionality. Analyses of personal identity admit of as much succinctness as possible. It is F that is required for and guarantees that  $x$  is  $y$ . While F can refer to a set of properties, rather than just a single property, ultimately it is the definitive variable that is alluded to in understanding of personal identity. With multidimensionality, it is made difficult to conceptualize specifics of what ground identity: multidimensionality grounds identity through the ability of an identity to be conceived of as taking on many different aspects and parts. Yet, in regards to contemporary analyses of personal identity the emphasis is upon noting specific particulars that establish personal identity, not a large ensemble of disparate characteristics.

There are two ways of responding to this. The first is that the authors inform their concept of identity by pulling from Shoemaker and Parfit in conceiving of it, via theories of psychological continuity. The authors state that, "[The] claim [that identity extends via memory] has been developed into a more elaborate theory in which identity is conceptualized as a product of psychological continuity and connectedness, which in turn is a product of links between the self in the past and the self in the present. Autobiographical memory has a clear role in the constitution of identity, enabling connections between discrete moments in consciousness, and, if lost, it would follow that one's identity would also be lost or significantly changed" (58), while citing psychological theories of identity. Not only this, but they also reference other prominent philosophers in attempting to understand the relation between autobiographical memories and identity. While the authors may not be explicitly talking about personal identity proper, their conception of identity is heavily informed by the literature of personal identity. And even if the authors can be shown to not be talking about personal identity proper, it should be noted that many others agree with Addis and Tippet in the conclusion that it is autobiographical memory that is most relevant in discussions of personal

identity. For example, Wilson and Lenart note the parallel of the role of memory for both personal identity and discussions of the extended mind. In following other theorists about memory's specific role in personal identity, the authors claim that, "it is narrative or autobiographical memory that is most directly relevant to discussions of personal identity. The sense of having a continued psychological existence over time, such that one can remember oneself having done certain things in the past, matters to us and is what allows us to guide our current actions and plan our futures in light of who we are" (Wilson and Lenart 431).

The second is that, even if the authors are not talking about personal identity proper, they are still undertaking an investigation to find out whether aspects of 'identity' are compromised because of memory loss. If their hypothesis is true that loss of memory leads to subsequent changes in one's sense of identity, we can at least use this in our investigation of personal identity and personhood under the notion that memory is important in grounding the concepts and ensuring the requisite persistence conditions in some way.

Recall that Addis and Tippett argue that, "if autobiographical memory does contribute to identity as proposed, it follows that a loss of this memory will affect one's identity" (58). Thus, the authors believe that the Alzheimer disease (AD) population presents an opportunity to study directly the impact of changes in both personal semantic and personal incident memory on identity (59). In their study, the authors administered two tests of autobiographical memory and two tests of identity to a group of mild-to-moderate AD subjects and to closely matched elderly control subjects. The autobiographical memory tasks, the autobiographical memory interview (AMI) and an autobiographical fluency task, provided measures of personal incident and personal semantic memory in three different lifetime periods. The AMI (Kopelman et al. 1990) was a structured interview used to assess participants' recall of personal facts over three lifetime periods: childhood, early adult life and recent adult life. Participants were asked to

recall three personal incidents from each of the three lifetime periods. The autobiographical fluency task (Dritschel et al. 1992) assessed the ability to generate both personal semantic and personal incident memories, as evaluated in equal sections for the two forms of biographical declarative memory. The personal semantic fluency section required participants to produce as many examples as possible of names of people known to them in a 90-second period. The personal incident fluency section required participants to produce as many personally experienced events as possible in a 90-second period. No great detail was required for each reported event; only a brief description so as to distinguish it from other reported events.

The identity tasks used the twenty statements test (Kuhn and McPartland 1954) and the Tennessee self-concept scale (Fitts and Warren 1996). These tests allowed the experimenters to derive measures of the strength, quality, complexity and direction of identity (62). Specifically, on the twenty statements test the strength of identity was assessed by the number of responses generated and by measures of the quality of identity (by the abstractness of responses) and complexity of identity (by the number of categories and subcategories sampled in the responses).

The twenty statements test required participants to give 20 statements in response to the question, "who am I?". Responses were classified into four categories of identity: attributes, social identities, evaluative descriptions, and physical descriptions. The following scores were calculated: the total number of responses (a measure of strength of identity); the percentage of responses that were coded as abstract (a measure of quality of identity); the number of categories and subcategories sampled in the responses of an individual (measures of complexity of identity).

The Tennessee self-concept scale measured three components of the self-concept (identity, satisfaction and behavior) over five domains (personal, family, social, moral and physical) (64). The test consisted of 82 descriptive statements which are rated for self-descriptiveness on a 5 point true-false scale (mostly true, partly true, neither, partly false, mostly false). It is with these tests that the authors sought to evaluate the relationship between autobiographical memory and identity.

The results of the autobiographical tests illustrated that the AD group generally performed more poorly than the control group (63). Generally, the control group scored better in both the mean autobiographical incidents score and the mean personal semantic score of the AMI. The results are consistent with a temporal gradient of impairment with no benefit of recency (i.e., poorer recall for autobiographical memories compared with remote autobiographical memories) (64). The same results were identified in the autobiographical fluency test. Interestingly, the AD group scored lower than the control group (showed lower autobiographical fluency), but consistently scored lower for each category. Both the AD and the control group consistently scored higher for the mean number of names recalled in early childhood than they scored for early and recent adulthood, albeit the control group was able to recall almost four times as efficiently in every category. The results for the identity test followed what was predicted based on the autobiographical memory tests. In the identity test, there was a significant difference between the AD and control groups, with AD participants producing significantly fewer responses than control group participants. One interesting thing to note is that, although the AD group had generally lower scores, the pattern of scores across the five subscales appears similar in the two groups. The scores of the AD group were lower overall than the control group, but there was no significant difference in the pattern of sub scores between

the two groups. The AD participants showed impairment of the measure of their identity but equally across domains.

In investigating the authors' prediction that impairment of autobiographical memory would result in changes in identity, regression analyses were undertaken in two ways. First, partial correlations followed by hierarchical regression analyses were carried out<sup>17</sup>. The results indicated substantive correlation between these concepts, with lower scores in the autobiographical section allowing the consistent prediction of lower scores on the identity tasks. The results of these regression analyses show that the impairment of personal semantic and personal incident memory in the AD group are most relevant for investigating whether impairment of autobiographical memory is associated with changes in aspects of identity. If we grant the validity of the regression analyses conducted by the authors, then it appears that

---

<sup>17</sup> Partial correlations were calculated between the performance of the AD group on tests of autobiographical memory and components of identity, controlling for global cognitive decline as assessed by the mini-mental state examination (MMSE) scores. Partial correlations controlling for MMSE score revealed significant correlations between the measure of strength of identity (total number of responses on the twenty statements test) and autobiographical fluency for early adulthood names, and autobiographical fluency for childhood events. Quality of identity, as measured by the percentage of abstract responses on the twenty statements test, was significantly negatively correlated with one measure of autobiographical memory. Largely the authors found that the poorer an individual's performance on both test measures of childhood personal semantic memory and one measure of early adulthood personal semantic memory, the lower the number of definite responses made on the Tennessee self-concept scale (66).

The second component of the regression analyses involved hierarchical regression analyses on the data of the AD group, conducted to investigate whether scores on identity variables were directly influenced by associated components of autobiographical memory. Each regression analysis sought to control initially for global cognitive decline by entering MMSE in the first block. The second block of variables entered comprised the components of autobiographical memory significantly associated with the specific identity variable, followed by a third block of variables consisting of the remaining autobiographical memory variables. Essentially, through the construction of regression models, the regression model comprising MMSE and a constant (block 1) did not produce a significant R<sup>2</sup> change over a model containing only a constant. The addition of autobiographical fluency for early adulthood names and autobiographical fluency for childhood events in block 2 resulted in a significant improvement in the model predicting strength of identity. As clarification, R<sup>2</sup> is a statistical measure of how close the data are to the fitted regression line. It is the proportion of the variance in the dependent variable that is predictable from the independent variables. Essentially, in constructing models to predict the variance of strength of identity scores, the addition of the results for the autobiographical memory tests significantly aided in the model's ability to predict strength of identity.

deficits in autobiographical memory result in changes in identity. Addis and Tippett note that this result is consistent with both reports of caregivers and qualitative studies which report a loss of identity in AD patients (69). While the authors do acknowledge that the relationship between impairment of autobiographical memory and changes in identity is not one-to-one, they do note that their results do indicate that changes in some measures of autobiographical memory in this AD group influenced changes in some aspects of their identity. The experimental findings of these authors provide empirical support for the claim that loss of autobiographical memory can lead to changes to identity.

The results arrived at by these authors may invite skepticism regarding their methodology. One area in particular is in the experimenters' evaluation of strength of identity in the twenty statements test. It is unclear that a neurotypical individual would receive a score as high as the control group consistently. For instance, scores were calculated, in one way, by the percentage of scores that were coded as abstract. Upon what measure could a response be coded as abstract? If an individual took the test and said that they were a philosopher, would the examiners take this to be an abstract answer considering that the participant omitted their history as a professor in philosophy?

Another aspect of the study subject to criticism is that the authors inadvertently assumed the conclusion in their experimental paradigms. If the twenty statements test was in part a measure of one of the categories of identity, namely social identities, then is it not the case that the results for this particular branch of identity would mirror some of the results in the AMI? It seems that some conclusions about strength of identity are inferred by a lack of memory. The authors are trying to show a correlation between these two concepts, that loss of biographical identity entails changes to identity. However, it seems possible that the authors may have inadvertently posed questions that can be construed as providing two positive (or

negative) analyses of both identity and autobiographical memory. This being said, while there is the possibility of some overlap, the authors did try to create controlled tests to individually assess each variable independently of the other. If there is the conclusion that one cannot test one without inadvertently invoking the other, then perhaps we either are not set in our methodology *or* it is because there is a significantly interesting interrelationship between these two variables.

The experimental findings by Addis and Tippet give further credence to the notion that memory, specifically autobiographical memory, is important to personal identity and personhood. Even if we grant that the study by Addis and Tippet is not enough to conclusively assert that persistence conditions are answered by appeals to memory, there is at least room to believe that memory plays a role in the articulation of personal identity, considering the work undertaken by these authors and the extensive history in personal identity that makes reference to the role of memory.

Analyses of personal identity allow us to identify an entity over time. They allow us to differentiate between two individuals, or even note when a person at t1 is no longer the same person at t2. Many analyses of personal identity posit autobiographical memory as a criterion for understanding the persistence conditions of a person; for understanding if their personal identity holds over time, whether they are the same person, one can look at their retention of autobiographical memory. Degradation to someone's autobiographical memories thus inhibits their ability to perceive themselves as a person with a detailed past. And with this degradation comes a degradation in their concept of who they are. The very person that they conceive of themselves as being is eroded with the loss of autobiographical memories.

This is where a difference between the authors' findings and our own conclusions can be drawn. The authors note how the loss of autobiographical memory, of memories of what happened during their lives, affects identity. Autobiographical memory loss leads to a patient's inability to come up with facts about themselves and their lives, illustrating a degradation to their self concept. With our understanding of personal identity and personhood, this study illustrates to us that these AD patients have troubles conceiving of their personhood. Likewise, this study illustrates to us that their personal identities have been altered by autobiographical memory loss, in that they are incapable of resonating and associating with certain lost autobiographical memories, memories that may have been very important in grounding their identity over time. This loss of autobiographical memories may prompt some to say that some AD participants are no longer the persons they used to be. Perhaps, on some analyses, this may be true. Certainly, however, the results of this study point us toward the conclusion that autobiographical memory loss impacts upon our conception of diachronic conditions of persistence, as well as our own ability to conceive of synchronic elements as surviving. If what we are as persons are our memories, or at least that memories help ground our specific identity as a person, then the loss of these memories will lead to a subsequent degradation of the person. The way we conceptualize the person will change with losses to their autobiographical memories, and the way that we conceive of these entities as persons may change. The results of the Addis and Tippett experiment support this sort of intuition: it is because these AD participants face a lack of autobiographical memory that the synchronic elements are degraded. And when the person is degraded, their ability to conceptualize themselves, their very concept of self, is harmed.

Recall that, if one accepts Clark and Chalmers' initial argument for the extended mind thesis, then one can acknowledge that Otto's beliefs can be extended beyond the body. Many

authors, Clark and Chalmers included, believe that this is enough to infer that the notebook acts as a repository for Otto's memories as well as beliefs. If we allow that memories play an important role in grounding the persistence conditions of personal identity, then it seems that it is at least feasible that these extended memories play a role in one's personal identity. If we allow this, then we are allowed the possibility that personal identity and personhood is best conceived of as an extendable kind. What follows is a discussion of how we can conceive of memory as extended, and how this might contribute to our understanding of personal identity and personhood as extended.

#### 4.1. Extended Memory

If persistence conditions of personal identity are conceived of through the role of memory, and if the conclusion about Otto using an external source for memories is correct, then in some cases the extension of the memories that are important for persistence conditions can be extended. Otto relies on and utilizes his notebook as effectively as Inga uses her internal memory storage for finding their way to the museum. The kind of parity considerations in play here can be used to motivate a broader rethinking of the kind of memory central to personal identity. However, there are areas to push back against this claim. For it is not the case that *all* memories are relevant to persistence conditions; some memories do not affect our judgement of whether a person has persisted from  $t_1$  to  $t_2$ . For instance, while it may be an authentic memory that you 'lived in apartment 32b, next to Shelton', this knowledge may not be relevant in understanding the persistence conditions of your personal identity. Whether or not you possess this memory at  $t_2$  will have no bearing upon whether you persisted. The point of this is that not all memories are necessarily considered relevant to our conception of persistence conditions of personal identity.

Klein and Nichols themselves largely dismiss the role of procedural memory for personal identity, claiming that the literature has largely focused on declarative memory in making sense of memory. On top of this, most authors that discuss the role of memory in personal identity specifically refer to autobiographical memory. Recall that Locke grounds the criterion of persistence in the ability to consciously track past experiences, actions, and thoughts. Identity can be guaranteed and reinforced in an individual as far as their consciousness can reach backwards. Addis and Tippett focus exclusively on the relationship between autobiographical memory and its retention, rather than focusing more broadly on the ability of individuals to utilize systems of memory generally. Most authors focus on declarative memory in understanding identity, just as Wilson and Lennart focus on the subtype of declarative memory (semantic memory), in that, “such narrative or autobiographical memory, particularly in its individualistic guise, might be thought to fall under the broad umbrella of declarative (vs procedural) memory since it is, more specifically, a type of episodic memory involving auto-noetic (or self-knowing) awareness” (Tulving 2010).

As such, it seems that even if Otto and his notebook meet all of the glue and trust conditions, and as such allow that the notebook and Otto are a both part of a system that establishes extended memory, we can still reject the conclusion that Otto’s person and personal identity extends into the notebook. If autobiographical memories, specifically declarative type autobiographical memories, are the only type of memory that is relevant to diachronic and synchronic considerations, then even the capacity for the extension of memories does not entail that extended memories are relevant in considerations of personal identity and personhood.

This conclusion will be challenged in the following two ways. The first is to challenge the assumption that only declarative autobiographical memories are relevant for personal identity. The second is to instead focus on how autobiographical memories can be considered extended

under certain circumstances. This will be done through an analysis of socially extended memories and certain cognitive artifacts and externals that facilitate the extension and retention of memory.

#### 4.2. Non-Autobiographical Memory

In discussing the role of non-autobiographical, and in particular, procedural memory for personal identity, I would like to discuss Schechtman's conclusions in her article "The Brain/Body Problem" (1997). Schechtman discusses the contemporary assumption that the mind is located in the brain. While it is claimed that this view is empirical, Schechtman will argue against this claim in favor of a distributed view of the mind. While there are noticeable correlations between mental activity and brain activity, evidence reveals that it is virtually never the brain alone which is active in mental life (Schechtman 1997). The brain and the body are in constant interaction, and the physiological understanding of mental functions almost inevitably involves not only an understanding of what is happening in the brain, but of the entire feedback loop between the brain and peripheral systems. As such, Schechtman advocates an alternative view that sees the mind, instead of as the sole locus of mental activity, as a central organ in a distributed system. "This view [the distributed view] conceives of the brain's role in producing mental phenomena as analogous to the role of the heart in circulation, or the lungs in respiration. The heart is, without a doubt, the center of the circulatory system, but it is not the place where circulation is located nor do we say that the lungs are the location of respiration. Circulation and respiration are distributed systems of the entire organism, with a major organ at their core" (152). The various mental features that make up the mind—cognition, emotion, sensation, etc.—involve a broad range of physical activity, and thus can be viewed as distributed systems of a human being rather than the activity of a single organ.

One example that she gives for thinking this is through the discussion of brain transplant cases. Schechtman notes that no real challenge is raised to the assumption that the transplant of a brain would result in the transfer of psychological features. Importantly, Schechtman points out that this belief in the transfer of one's personal identity to another body is only done through the assumption that the brain serves as the repository of the psychological subject. It is this assumption that allows the belief that the brain is enough to transplant all of the beliefs, desires, memories, and character traits of a person. Schechtman claims that there is little empirical justification for expecting this outcome. "Brain transplant cases assume that a person's memories and personality would be moved with her brain, but a bit of reflection on common knowledge about the physiology of psychology makes it doubtful that either would be straightforwardly transferred" (158). Schechtman considers memories, and that it is less likely that all memories are most plausibly stored in the brain alone. Memories that have to do with motor skills (procedural memory systems) seem to crucially involve the rest of the body. Schechtman discusses the example of a pianist, who has practiced a piece repeatedly until she knows how to play it. Schechtman claims that this pianist importantly *remembers* the piece. Schechtman notes that, "it seems unreasonable to believe that this memory is captured exclusively by what goes on in the brain or that we should predict that if the pianist's brain were transplanted into any body with functioning fingers that that body would be able to play the piece. Whatever is involved in the memory is at least partly in her fingers, and in the coordination between brain and body, and the parameters of this coordination are at least part of what is stored in the brain. It therefore seems reasonable to suppose that this memory and memories like it would be lost in a brain-transfer" (158).

I believe that this example is especially relevant to our understanding of memories that are important to personal identity. The above example portrayed the case of a pianist's brain

being transplanted into another body and not carrying with it its procedural memories of piano. We could imagine a virtuoso pianist who understands her personhood largely through the framework of her involvement in the world of music and pianos. Upon discovering that she no longer has access to this ability that defines her she may conclude that she was subject to a medical invasion, and that this transplant has irreparably taken a part of her away. Schechtman is of the same opinion, stating that, "certainly one's virtuoso piano skill or sprightly step can be essential to one's personality as many experience memories" (159). If we consider procedural memory as the system responsible for the acquisition and retention of motor, perceptual and cognitive skills, and if it is the case that this discussed ability is a procedural memory, then we can conclude that in certain problem cases the failed retention or changes to certain procedural memories results in changes in personal identity and personhood. In this sense we can conceive that it is not just autobiographical memories but also other memory systems that may be relevant to our conception of personal identity and personhood.

Not only are procedural memories relevant, as revealed by the above discussion. We could imagine that semantic memory that is not self-referential may also have a role in the articulation of personal identity: say one subscribes to something of a narrative analysis of personal identity, understanding their life as a person to be informed by the aspect that they are a well renowned scientist. They pride themselves upon their intellectual mastery of different scientific domains. Such knowledge consists of base semantic memory of different algorithms and route memorization, among other things this thesis on philosophy is glossing. Their identity and conception of themselves is based upon the fact that they hold this information. Say this individual were to face a disruption to their memory, such that their practical memory of scientific knowledge was lost. This scientist may face a crisis of identity, as he may think that the narrative structure by which he understood his life and person is compromised. It may be that

the volitional structure by which they orient their life is fundamentally changed. The post-memory individual may concern themselves with reconstructing their knowledge of science so as to reclaim their 'identity'. Or perhaps they decide to concern their life now with the import of their families or community, instead focusing on their actualization as a parental figure, thus framing their identity around that. If we take either a volitional view of personhood or a narrative view of personal identity, then we could infer that the loss or change to non-autobiographical semantic memory can impact upon personal identity and even the status of personhood.

Ultimately, this argument only bears on authors who are willing to allow that other forms of memory may be relevant in personal identity. The claim may be made that certain procedural memories are necessary for the actualization of personal identity and personhood, but largely do not have any bearing on these concepts. Sure, it is important that an individual's procedural motor memories play a role in grounding the capacities for other autobiographical memory systems, but this does not mean that significant changes to *all* forms of procedural memories are relevant. We do not necessarily consider a stroke victim that loses their capacity for language as losing their personal identity or personhood. Theorists that hold to the importance of the criterion of memory may hold to the notion that it is primarily autobiographical memories that are relevant in their analyses of personal identity and personhood. However, it seems that there is at least room to problematize this stringency by some, as elucidated above. When we recognize the importance of memory systems, other than that of autobiographical memory, to the concepts of personal identity and personhood, we are allowed some leniency in understanding the possible extension of Otto's personhood and personal identity. Since we have denied that only autobiographical memories are relevant to personal identity and personhood, we can allow the reasoning that Otto can be considered as

extending even if his notebook does not, or cannot, act as a part of an extended system for autobiographical memory.

The next section focusses on how we can conceive of autobiographical memories as extended under certain circumstances, and how we might imagine Otto's notebook as acting as part of an extended autobiographical memory system.

#### 4.3. Autobiographical Extended Memories

There are other methods we can utilize in thinking of autobiographical memories as extended. One way is through looking at transactive memory systems and collective memory. Heersmink (2017) notes how cognitive systems are often scaffolded and sometimes constituted by external resources. This is in part because situated cognition theory has pointed out that our embodied brains are powerful pattern recognition and completion systems, but have limited information-storage and processing capacities (3137). In arguing over whether personal identity and the self can be partly constituted by external informational structures, Heersmink notes that the literature on transactive memory reveals instances of cognitive interdependence between agents (see Theniner 2013). Specifically, Heersmink notes how other authors have emphasized the interdependence of individual memory system and therefore point out that transactive memory systems cannot be reduced to individual memory: it is an emergent group-level property that exists only when people interact and communicate in the right way (Heersmink 3137). These transactive memory systems, like other memory systems, cannot be reduced to their individual members. In this way we can understand memory as properly articulated, in certain relevant circumstances, as being properly understood when extended. Ultimately, Heersmink will claim that, "if memory is often distributed and if the self is partly constituted by our memory, then the self is also distributed" (3136). As such, we can understand his

articulation of transactive memory as advancing a view of memory where it is not fully conceptualized in specific circumstances when we look at individuals alone.

Largely, the conclusion drawn by the discussion of transactive memory is that memory need not be bound exclusively inside the skin-skull boundary, since it may turn out that “individuals” extend. In pursuing the thesis of how narrative memory need not be bound to the skin-and-skull boundary but can incorporate the world beyond the body in a variety of ways, several authors note the effect of cognitive offloading for memory. On top of transactive memory systems, which use collaborating and relying on other people in performing memory tasks, Heersmink argues that we scaffold our memory with various other external artifacts, referencing how bartenders and cooks offload the chore of remembering drink orders onto their environment, arguing thus that material culture plays an important role in human memory. Wilson and Lenart (2015) note how proponents of extended cognition take such examples of cognitive offloading to point to how extended cognitive systems have been shaped evolutionarily, developmentally, and culturally for everyday cognitive tasks (430). Daniel Dennett (1996) illustrates our habit of offloading cognitive tasks into the environment with the example of AD and the elderly who have difficulties recalling daily routines and suffer from other memory-related deficiencies once they are housed in nursing homes and other palliative care institutions. Dennett notes that many signs of dementia are less pronounced or disappear altogether once people are returned to their own homes where they have offloaded many of their daily routine schedules on items or places that remind them of their obligations, how they ought to pursue them, and other kinds of pertinent information. In such a way we can see how individuals can offload their memory upon familiar environments, and in such a way actually sustain their self-conception and personal identity.

Wilson and Lenart note that, as with the above example, integrating things in one's immediate environment to form an extended memory system is a form of extended cognition used by people with AD and other neurodegenerative conditions affecting memory; "here is the individual who remembers, but the activity of remembering is extended, being distributed between that individual and things in her environment" (432). Dennett notes that, "taking [AD patients] out of their homes is literally separating them from large parts of their minds (Dennett 138). Relevant externals help provide a self-structured and stable environment that provides continuity for their memory where their individual faculties otherwise cannot function optimally. If we endorse the experimental findings of Addis and Tippett, or at the least accept a psychological analysis of personal identity and personhood, then it seems that elderly individuals who offload onto their environment are able to maintain their cognitive faculties and provide continuity for their ailing faculties of memory. Granting that these individuals are providing continuity for their memories, and that it is possible that the memories whose continuity is established are autobiographical in nature, then it seems tenable that these individuals have autobiographical memories distributed upon their environments, and thus can be seen as having a form of extended identity.

Wilson and Lenart pursue this thesis, specifically noting that narrative (autobiographical) memory can be perceived as incorporating the world beyond the individual in a variety of ways, doing so by addressing the social dimension of thinking of extended/distributed memory in terms of the relationship between autobiographical and collective memory. The authors believe that narrative and autobiographical memory can be shared and co-constructed by two or more individuals (431). By 'collective memory', they refer to the commemorative, ritualistic and political retention of significant past events. Wilson and Lenart reference the way we collectively remember atrocities on the calendar date on which they were committed, or how we engage in

joint actions that express our political affinity and sympathy with (other) victims of a crime or natural disaster. The authors believe that they can maintain a version of the distinction between narrative and collective memory by thinking of the sharing of one's narrative memories in the same way that we can think about the offloading of those memories. "The characterization of collective memory can challenge the notion that it is individuals that remember, but rather systems. In regards to the cases of collective memory, it is some kind of collective or group that remembers, distributing the task of remembering between different individuals within a group in ways that make it implausible to identify any one of them as "the" person who remembers" (432). If we allow that autobiographical memory can be like transactive and collective memory in such a way, then it seems that we can look not just towards our environment but also towards the individuals and social groups within that environment for grounding our personal identity.

With all of the above shown, Wilson and Lenart note that, "when we take memory seriously in the context of personal identity, it becomes clear that individual identities, just like individual memories, are realized within the context of collective narratives" (432). Individual memories may well serve as the vehicles for individual identities. But such memories are influenced by collective narratives, thus making individual identities heavily reliant on the collective or social contexts within which individuals exist. In this same way, other authors similarly conceive of the relationship between persons and their environments. Schechtman conceives of persons being what they are because they were encultured and raised in a person-space, and are characterized by actualizing person lives within such a space. It is their environment that is ultimately necessary for the actualization, recognition and existence of the individual person, just as the individual person contributes to and sustains the person-space (Schechtman 2014, 113). With the above postulated, Wilson and Lenart claim that, this kind of

relationship between individual narrative and collective remembering suggests that the extended mind thesis may be well positioned to augment traditional neo-Lockean views of personal identity. Supposing that the cognitive capacities involved in remembering are not intrinsic to the individual whose identity is being tracked, but are socially and environmentally manifested capacities, then this would imply that personal identity has a wide realization.

Recall that Heersmink claims that we scaffold our memory with various external artifacts, and that material culture plays an important role in human memory. Specifically, Heersmink investigates how situated autobiographical memory technologies (referred to as 'lifelogging technologies', technologies designed to record aspects of one's everyday life, such as fitbits or sensecams) have beneficial effects on a patient's autobiographical memory such as maintaining the integrity, delaying the disintegration, or in some cases replacing autobiographical memory.

If we allow that transactive memory systems, the scaffolding of memory onto artifacts and the utilization of autobiographical memory devices that aid in our ability to retain and maintain the integrity of memories, and if we allow that memories can be seen as an important criterion for personal identity, then it seems clear that cognitive artifacts and appropriate externals can be seen in certain situations as part of token instantiations of personal identity. Recall the Lockean conception of personal identity. It is consciousness that entails the attribution of personhood. But it is psychological connectedness of  $x$  and  $y$  that guarantees shared personal identity between them. "For as far as any intelligent being can repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present action; so far it is the same personal self" (Locke 40). The points made by Heersmink guarantee at least that cognitive artifacts and externals can aid in the

retention and maintaining of memories. Thus, these externals help ground personal identity, when read under the criterion of memory.

Heersmink moves on to characterize not just our memory, but our cognitive capacities as important in characterizing our personal identity and personhood. He argues that certain cognitive capacities are important for the kind of person we are, and thus that cognitive artifacts that are used in processes of offloading and extending are thus important for a conception of personal identity. This is in part due to the fact that our cognitive capacities are relational in that they often depend on social and artifactual structures in the environment in which we are situated (Heersmink 3143). Thus, if we grant that external resources are essential for our cognitive capacities, and that these capacities are important for our conception of diachronic and synchronic aspects, then we can infer that these external resources are not only incredibly important in our construction of personal identity and notions of personhood, but also that we are better suited at characterizing these externals as things our personal identity and personhood extend into. Not only does Heersmink think that cognitive artifacts can have an important role in understanding conceptions of personal identity, but he also believes that these artifacts have a moral status in some cases. Heersmink (2015) claims that ethicists have largely focused on the moral implications of cognitive enhancement through medical or therapeutic purposes, focusing on the implications of psychopharmaceuticals and how the enhancement of cognition through such means can compromise moral values like safety, freedom, authenticity, fairness, etc. (Heersmink 21). Heersmink instead chooses to focus on the moral aspects of enhancing cognitive abilities with the aid of external artifacts (tools such as notebooks, checklists, calculators, etc), a domain that he claims is neglected in the cognitive enhancement debate.

The neuroethical discussions of cognitive enhancement have ignored the role of external artifacts. Yet this being said Heersmink does not think that these artifacts themselves are not without moral status and implications. If one agrees that tampering with information in Otto's notebook is akin to tampering with information stored in his brain, and that tampering with an individual's brain can be problematic, "as our identity and capabilities are largely determined by the brain" (Heersmink 27), then one can infer that external information, in some cases, has moral status. This follows from the claim by Clark and Chalmers that, "in some cases interfering with someone's environment will have the same moral significance as interfering with their person" (18). If one allow that these external artifacts and external environments (that allow for extended cognitive systems and minds) have moral value, then we have room to deduce that these artifacts and environments are morally important in the sense that they are valuable to one's personal identity and the moral questions that surround such questions. With the recognition of Otto's notebook as having moral status the further step can be taken that its moral status is also essential in understanding certain harms done to persons and affronts to personal identity. Finally, if all this is allowed and we recognize that some external artifacts can play important roles in understanding the moral parameters to affronts to personal identity and personhood, we can simultaneously infer that these artifacts are subsequently important to the very construction of the personal identity of certain extended individuals and should be seen as important embedded/extended parts of that individual's personal identity.

Wilson and Lenart argue that this approach to viewing memory, and personal identity and personhood, as beyond the skin-and-skull boundary allows us the resources to solve some of the problems facing standard Neo-Lockean views, such as problems that stem from the individualistic ratio-centrism of such views. For instance, there is a notion that a certain level of rationality is necessary to be a person, and thus the status of fragile individuals and of other

entities is threatened by this conception (Lenart 2004). This understanding of the wider base of realization for persons and memories is done by recognizing that narrative-based criteria for personhood are based on more than just the intrinsic cognitive capacities underpinning the remembering of those who are neurotypical, as well as allowing the psychologically based criteria to have a wider base of realization (434). These notions are also supported by Lindeman (2010) who argues that a person's identity is both shaped and preserved by others via the complex interactions between, and varied intertwining narratives remembered and transmitted within, families and other groups. "It's not just other people who hold us in our identities. Familiar places and things, beloved objects, pets, cherished rituals, one's own bed or favorite shirt, can and do help us maintain our sense of self" (Lindeman 162). In this sort of way, an externalist approach affects how we define and track personhood, as well as change the way we understand the mind (Wilson and Lenart 434).

Individuals with cognitive limitations that create difficulties for tracking their own identities, thus magnifying the problems that we all face in preserving a coherent conception of ourselves, may depend on others more deeply to maintain such cohesive narratives. Consider fragile individuals, persons who slip gradually, over time, into ongoing states of dementia (Lenart 2014). On an externalist account of personhood, such persons need not forfeit their identities as their minds and memories deteriorate. This is because even though they gradually lose their capacity for individual memory, their identities are realized in the collective remembering of others, as well as supported and externalized by their environments. Wilson and Lenart thus posit that, "the externalist account of personal identity thus reveals a fragile side of personhood that remains hidden in individualistic variants of the Neo-Lockean approach" (435) Understanding the sensitivity of narratives to interpretive intervention deepens our

understanding of what it means to treat people and their identities authentically, allowing us the insight that personal identity is maintained by both individual and collective manifestations.

In conclusion to the above arguments for the extension of memory, specifically autobiographical or narrative memory, as being illustrative of the tenability of the extension of personal identity and personhood, I would just like to note some of the following. The above authors focus almost exclusively on the aforementioned possibility of the extension of personal identity. While there is discussion of the extension of personhood, it is done fleetingly. I intend to discuss what I consider to be a relevant issue for the tenability of extended persons, and as such will provide further analysis of what the extension of this concept might look like. This being said, many of the intuitions and arguments addressed above translate over to discussions of personhood. For instance, recall the argument that degrees of offloading upon one's environment can help fragile individuals maintain their mental capacities and identity. If personal autonomy is a measure by which we gauge the presence of personhood, then one could argue that certain fragile individuals can be placed in precarious positions where their status as persons are jeopardized (for instance, issues of consent might arise because of a perceived deficiency in personhood). If we grant that one can retain their status as a self-functioning cognitive agent by extending upon their environment, then there is room to argue that in some cases fragile individuals maintain their personhood through externals.

## 5. Conclusion

The purpose of this chapter, and largely of this thesis, is to open up considerations of the extension of personal identity and personhood: it is possible to conceive of personal identity and personhood as extended. Recall, however, the introductory claims about the verification of extension. We cannot prove that personal identity or personhood can extend unless we are

willing to commit to a particular analysis of personal identity and personhood. Proving the extension of these concepts will require more work than will be done in this thesis, primarily that of arguing for a specific analysis of the concepts and showing its superiority, then illustrating how this analysis is compatible with extension. As it stands, this thesis has argued that the reasons given for rejecting the possibility of extended personal identity and personhood are not strong enough to establish their impossibility. Granting also the work done in this chapter in recognizing the connection between memory and personal identity, we can accurately understand personal identity and personhood as extended. While authors like Milojevic and Wilson and Lenart can claim to verify the extension of personal identity, this thesis instead has so far focused on clarifying the foundations upon which EST and claims about extension stand. Through defending and strengthening HEC and EST, this thesis has provided a foundation to allow all analyses of personal identity and personhood to consider extension. Through investigating some ways in which the methodological conceptions of personal identity and personhood can create problems for the implications of extension, I have shown that personal identity and personhood are not incompatible with extension.

The following chapter provides some insight into how various contemporary theories of personal identity and personhood can be read through the lens of extension. The reading of these analyses will be strengthened by analyses given by other authors that support more general ideas of extension. Ultimately, the purpose of the next chapter is to show how the work done throughout this thesis can pay off in a meaningful way when we analyze contemporary theories of personal identity and personhood.

## Chapter 5. Extension of Personal Identity and Personhood

### 1. Introduction

Recall that, with the last chapter, the conclusion was reached that the extension of personal identity and personhood can only be verified through the initial positing of an analysis of these concepts, and then arguing for their subsequent extension. With this said, an analysis was undertaken to show one way we might conceive of personal identity and personhood as extended. As we saw in the last chapter, one way to conceive of extension is through an analysis of the role of memory in grounding personal identity and personhood. This chapter will aid in furthering the purpose of this thesis, namely to broaden the possibilities of seeing personal identity and personhood as extended. As such, this chapter will endeavor to briefly discuss and analyze how some prominent contemporary theories of personal identity and personhood can be understood as being compatible with extension.

This chapter will analyze the views of a few prominent contemporary philosophers who contribute popular analyses of personal identity and personhood. What follows is a brief exposition of Marya Schechtman's conception of personal identity and personhood, as well as some analysis to support the idea that her account can accommodate extension. We will start by expounding her general account of personhood, then move on to an analysis of how she understands persistence conditions. We will then analyze several contemporary accounts of personal identity and personhood to assess the compatibility of extension. Finally, a discussion of a novel problem facing this thesis will be discussed, namely an extrapolation upon a discussion by Derek Parfit (1971).

## 2.1. The Person Life View

Schechtman articulates an analysis of personal identity and personhood which she refers to as the Person Life View (PLV)<sup>18</sup>. This view posits that personhood is entailed by living 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 2014 111).

Schechtman articulates her view in the form of the following biconditional: “S is a person if and only if S is an entity who lives a person life and continues for the duration of a single person life” (2010 278). A person life is the kind of life typically lived by an enculturated human. Thus a person is an entity that lives a person life. In understanding synchronic conditions for personhood, we first must understand what a person life is.

From here some clarification is needed, as Schechtman’s analysis of what entails personhood and person lives is understood by way of a schematic investigation. In characterizing person lives, Schechtman initially attempts to reflect on the most prevalent paradigmatic examples of person lives. She does this out of necessity, as she claims that, “any account of personhood and personal identity will need to start with paradigmatic cases on which there is general agreement and move from there, seeking to determine the relevant factors in the standard cases and offering a principled way of determining how much deviation is possible before personhood and personal identity are undermined” (2014 111). Thus Schechtman focuses on healthy, enculturated humans as the paradigm of personhood. She notes that these person lives begin with a period of social dependence, and relatively basic

---

<sup>18</sup> This thesis primarily relies on Schechtman’s book *Staying Alive* (2014) in expositing PLV. This being said, Schechtman has argued for this view in other places, and as such this thesis also makes use of her article “Personhood and the Practical” (2010).

cognitive capacities which develop over time. Importantly, Schechtman notes that the psychological and physical development of humans requires proper scaffolding and social support from the environment. Likewise, there are particular psychological capacities that are required if one is to engage in the more sophisticated kinds of interpersonal interactions found in a standard person life. Without the ability to develop these internal resources, active participation in a person life would be severely limited (112).

When thinking about the typical person, Schechtman notes that they are sentient, reflectively self-conscious, rational and moral. These and other psychological capacities are required if one is to engage in sophisticated kinds of interpersonal interactions found in a standard person life. Granting that individuals are not born with all of these capacities already developed, Schechtman notes the importance of maturing in an environment that provides the proper scaffolding and social support for such development.

Having established a basic conceptual criterion for persons, Schechtman goes on to conceptualize three main components of a person life. The first is the set of attributes an individual possesses, i.e., the physical and psychological capacities and internal structures that need to be possessed to be considered living a person life. The second component of a person life is the set of activities and interactions that make up the individual's life. An individual's life will involve particular relationships and activities, though for the purpose of understanding this second component the importance is not on the specifics of the relationships and activities. Rather it is whether the day-to-day activities and interactions of the individual are part of the general form of life of typical enculturated humans (113). The third element is what Schechtman calls the social and cultural infrastructure of personhood—"the set of practices and institutions that provides the backdrop within which the kinds of activities that make up the form of life of personhood become possible" (113). In conceptualizing person lives, we thus cannot exclusively

look at individuals to recognize the existence of a person life. We have to look not only at individual attributes and their social interactions and activities, but also at the stable background structures that make these things possible.

This third component requires a bit of explanation, as it plays a large role in a later articulation of person-space. This infrastructure will contain presuppositions about what gets brought into the form of life that is personhood. It is this infrastructure that informs us about what particular roles human persons are ascribed, independent of their individual attributes. This infrastructure supports the social and cultural attitudes we bear towards other human persons, specifically that when we encounter other humans, “we automatically see them as persons and interact with them as such” (113). Schechtman claims that our ‘default’ stance of taking this attitude toward particular kinds of others is institutionalized as part of our social and cultural infrastructure. The larger claim is that there will be some practices and institutions that are necessary for there to be person lives at all.

Person lives are enabled within person-spaces. More importantly, person lives cannot be lived without the existence of a person-space to facilitate the components necessary for them. A person-space is the social and cultural infrastructure within which persons interact and which supports personhood. A person-space is both physical and immaterial, in that the person-space enables the organization of human persons to live together<sup>19</sup>. It is within person-spaces that human functions are mediated and transformed. Person lives, in their articulation, are

---

<sup>19</sup> Schechtman notes that, “whatever the relation of culture to society, the kinds of organization they describe, interacting in whatever way they do, together make up what I am thinking of as person-space—the social and cultural infrastructure within which persons interact and which supports personhood. I will use “culture,” “social infrastructure,” “person-space,” and like terms more or less interchangeably to refer to this constellation” (115). Thus this constellation of things that constitute this infrastructure can include codified systems as well as unregulated practices that are shaped by culture, such as systems of entertainment and leisure.

dependent upon a person-space that gives us an understanding of how the particular capacities are actualized in person lives, as well as providing the backdrop of the cultural and societal infrastructure necessary for development. Likewise, the infrastructure of person-spaces survives symbiotically, in that, “the infrastructure could not exist unless those who are part of person-space have the necessary capacities and attributes to create and maintain it” (115). Person-space only exists because there are persons to maintain it, and persons can only exist and actualize when there is a person-space to provide necessary infrastructure for this actualization and existence. The claim by Schechtman is that, in order for there to be beings with our capacities, there must be a certain kind of social infrastructure in place, and so the capacities depend upon the social infrastructure just as much as the social infrastructure depends upon our capacities in maintaining and constructing person-spaces. “There would be no persons without person-space, and to be living a person life is to be accorded a place in person-space, to live as and be treated as a unified locus of the sorts of practical concerns and interactions that typify the lives of those who generate and maintain the social/cultural infrastructure within which these lives take place” (118).<sup>20</sup>

With the above exposition of PLV given, how can we go about evaluating the thesis of extension in relation to it? I believe that there is a myriad of ways to go about reading PLV in such a way that we can conceive of this view as being compatible with extension. Schechtman maintains that person-space is maintained by people in it, and that persons can only exist when there is a person-space to provide necessary infrastructure. This social and cultural infrastructure is incredibly important for our conception of personhood. Without this relevant infrastructure, there would be no persons.

---

<sup>20</sup> More details on PLV will emerge as we continue to investigate her view

This alone does not guarantee that personhood extends beyond human bodies however. Schechtman supports the assertion that facts about our embodiment as humans heavily inform how we understand personhood. Still, there is something to be said about the looseness by which we identify persons on this view. Personhood is something articulated only within person-spaces, and it is a concept maintained by the collective work of culture, institutions, individuals, physical facts about organisms, etc. A correct reading of Schechtman's analysis is that personhood is not something maintained by individuals alone. And as such, the individual ascription of personhood is not something that is maintained by the individual alone who it is attributed to. An individual's attributed status of personhood is not something that resides within the skin-and-skull boundary in this sense, but is instead maintained by many relevant external sources.

With all this said, the source of the attribution of personhood can be shown to be beyond the body. Yet this itself is not the same thing as the conclusion that persons extend beyond their bodies. Even with this granted, however, there is enough here to question the notion of a boundary in PLV. The mere fact that personhood is not something that can be maintained exclusively by the individual it is attributed to is enough to problematize how we conceive of the boundary established between person and body, in several ways. For instance, even though Schechtman focuses exclusively on human persons in her exposition of person lives and person-spaces, she allows that there may be non-human persons as long as they have the forensic capacities (e.g. capacities for language, acquisition of emotional states necessary to flourish in a human person-space, etc.) necessary to actualize as persons in a person-space. Thus while human embodiment plays an important role in the infrastructure that maintains person-spaces, human embodiment is not a necessary condition for personhood. She states that, "PLV denies that human persons are necessarily identical to human animals" (111). If we allow that a

non-human with a different form of embodiment could in principle be a person in a human constructed person-space, then why not think that personhood can be attributed to alternative forms of human embodiment?

Schechtman talks about the complication of attributing personhood to entities that do not follow a typical developmental trajectory. She states that, “it is true that these are not paradigmatic person lives, but these individuals are clearly seen as persons and given a place in person-space even when it is evident that they will lack some of the standard capacities persons usually possess and will not be able to participate in the full range of person-related activities” (120). Thus while the conception is that certain capacities are central to the articulation of personhood, what is important is that the person-space emphasizes these capacities in what is vital to be developed.

Infants and individuals who do not follow typical developmental trajectories are afforded a place in person-space irrespective of the actual presence of forensic capacities that Schechtman identifies as important for personhood. Schechtman justifies this by claiming that, “PLV sees humans with atypical developmental prognosis as persons for much the same reason that it sees human infants as persons—because there is a default expectation that such infants will develop into beings with the full complement of forensic capacities; an expectation which is over-ridden in the atypical cases but does not disappear or cease to do work even when we know that expectation will not be met” (122). Schechtman does not see this as overly complicating for her analysis of personhood, though there is the possibility for criticisms here.

The important point of interest for our investigation is that personhood is *attributed* to persons by virtue of their membership within a person-space. This person-space articulates the forensic capacities that are necessary for the living of a person life. Recall that Schechtman

importantly states that there are no persons without person-spaces. Thus while a set of forensic capacities important to a person-space may be present in an individual, ultimately if the individual is not part of this space then they are not a person. Likewise, the actual presence or not of forensic capacities does not matter for the attribution of personhood; personhood is maintained irrespective of these capacities. Thus, what maintains personhood and what allows us to attribute it to individuals is not determined by qualities of the individual alone, but by an understanding of the kind of life they are allotted in a person-space. The attribution and maintenance of personhood thus is extended beyond the individual. But what of persons themselves?

Here I will provide two reasons for thinking that persons can extend. For the first one, we need to recall that on the PLV, S is a person if and only if they live a person life. If so, then by what parameter can we gauge whether persons extend? We have shown, with previous claims, that personhood is not attributable to humans alone. Likewise we have noted that the status an individual bears as a person is maintained in part by infrastructure external to the individual. As such there is no basis for identifying the boundary for persons. This may provide an opportunity for us however. Since being a person entails that, “one live a person life, which means engaging with others in person-specific ways codified in the social infrastructure, and taking a place in person-space” (131), then this gives us a little bit of room to articulate what kinds of entities might be ascribed personhood. When talking about the relationship between the infrastructure of person-spaces and of persons, Schechtman states that, “the development of institutions and norms of the sort that define person lives (e.g., prisons or systems of punishment, economies, theologies, art) requires beings who have certain kinds of memory systems, reflective self-consciousness, rationality, and related cognitive and affective abilities—i.e., the forensic capacities. The notion of personhood that is inscribed in these institutions and norms will thus

be one that recognizes that persons must typically possess these capacities (for at least part of their lives) or they will lack stability and efficacy” (116). Schechtman claims that person-spaces are supported by beings with certain forensic capacities, and that these beings may live person lives in virtue of both their forensic capacities and the person-spaces they reside in, seems to leave room for the further claim that these forensic capacities may extend.

However, I think there is room to push back against this. Recall that PLV views some human individuals with atypical development as persons, even if they lack these forensic capacities. Their positions within person-space and the roles attributed to them are limited, and in some cases the loss of enough of these capacities can result in an almost completely diminished person. PLV can be seen as articulating the claim that, *almost* irrespective of forensic capacities, personhood can be held as long as the entity is perceived as a constituent member of a person-space. Regardless of the difficulties here I believe that we have good reason to think that some of the sufficient conditions of personhood may extend.

If it is only by virtue of living a person life that someone is a person, then perhaps we can allow that the grounds of someone’s abilities to live a person life can extend beyond the body. If it is the fact that an entity is a person through living a person life, and that a more actualized person life is indicative of a higher degree of personhood, then we could imagine that someone may be able to more highly actualize when their capacities to live a person life are expanded. Say for instance that Otto is part of a tennis club, but because of his condition he cannot remember the times and location of this club. Through his notebook he is able to keep up his membership with the club and is able to continue an important social activity that characterizes part of his lived person life. The notebook allows Otto to live a more actualized person life by expanding forensic capacities that help maintain his person life. Without his

notebook he would be less of a person and would lead a person life with less components fully met.

Let us imagine a version of Otto with severely limited capacities. They are so limited, in fact, that Otto's person life can be characterized as being almost devoid of one of the important components that characterizes a person life. In particular, Otto is incapable, on his own, of participating in important activities and interactions that previously made up his person life. We can imagine that this deficiency comes from a severely limited faculty of memory, where if he is not prompted by something external to himself he will only take care of his base subsistence. His notebook saves him from this existence by acting both as an external source of memory, informing him about appointments and social obligations he may have, and as something of a prompt to remind him about his desires to engage important activities and interactions. In this example, it seems that Otto cannot be characterized as living a person life on his own, but that through the hybridization with his notebook he is able to live a person life. In this sense it is best to conceive of the entity that is living the person life as the Otto-notebook hybrid, for without the notebook Otto is not living a fully actualized person life. Otto without his notebook cannot be characterized as living the same person life as Otto with his notebook. Importantly, remember that we individuate persons by individuating person lives. Thus it becomes necessary to acknowledge the notebook in individuating Otto from others. In this substantive way I believe that we have good reason to think of Otto's notebook as constituting the capacities needed to live a person life, and beyond that that the notebook constitutes the entity in that we cannot properly individuate and identify Otto without acknowledging the notebook.

Thus, in regards to the extension of synchronic elements, I believe we can conclude the following: because of Schechtman's claim that there may be non-human persons that live person lives, we can infer the base possibility that PLV can accommodate persons that extend.

Beyond this point, we have shown that we can argue that the forensic capacities of persons can also be shown to extend. Finally, we have shown that it is possible to imagine a fully lived person life only being made possible by the externals that constitute the components necessary for a person life. By this claim and the claim that individuation is done by individuating person lives, we can claim that in certain instances we can only identify individuals when we identify the relevant externals they are hybridized with.

Before moving on with our analysis, it will be important to quickly address how Schechtman's view answers questions about diachronic considerations. Schechtman proposes to apply Richard Boyd's concept of a homeostatic property cluster to our understanding of the diachronic conditions of person lives and persons. Boyd believes that the property-cluster conception of definition provides insight into forms of natural definitions (196). Boyd posits that a homeostatic property cluster is instantiated initially when a family of properties (F) are contingently clustered. The co-occurrences of these properties is the result of a homeostasis. "Either the presence of some of the properties in (F) tends to favor the presence of the others, or there are underlying mechanisms or processes which tend to maintain the presence of the properties in F, or both" (Boyd 197). Importantly it should be noted that there are theoretically/practically important effects which are produced by occurrences of properties (F) together with the underlying mechanisms and that, "imperfect homeostasis is possible and natural: some things displays some but not all of the properties of F" (197). Here arises the question of whether it is a normal feature of homeostatic property clusters to be predisposed to having some but not all properties of F act emergently, or whether this is an accidental feature of the homeostasis and that a correct and perfect one ought not to face this problem of some properties F not being displayed. For the sake of this essay it will be assumed that the accidental non-occurrence of properties F is a normal feature of homeostatic property clusters. From this

there is a kind term (t) which is applied to things in which the homeostatic clustering of most of the properties in F occurs. Thus it is all or part of the homeostatic cluster together with some or all mechanisms that underlie it that provide the definition of t (197). Importantly, the point is advanced that, “the kind of property denoted by t is a natural kind” (197). The homeostatic property cluster thus acts as harbinger and also is equally a conduit for the reality of t. Granting that the homeostatic property-cluster definition isolates real things within the natural world, Boyd utilizes this to advance his argument that, “naturalistic theories of reference and of definitions might be extended to the analysis of moral language” (201).

Schechtman supports an analysis of diachronic conditions through a cluster concept by appeal to an argument given by Olson (1997). Olson believes that we should be able to identify and reidentify biological life without an independent way to identify all the material parts which are caught up in it. Through an allusion to the ways we identify things like storms, he notes that even though storms involve material parts that are acted upon in particular ways, fixing the identity of storms and organisms does not require that we be able to know which material parts these are. What is necessary is being able to follow the series of activities that makes up an event of the appropriate kind (Olson 139). Schechtman believes that we can expand this so as to conceive of person lives and human persons as relevantly similar: “there is no presumption that there *is* any independent object, and so identity conditions are clearly set by the activities themselves” (147 author’s emphasis).

Some of the properties that can be identified as important for the property cluster are characteristic psychological capacities and engagements in certain kinds of social interactions, among other aspects detailed earlier. Included in this cluster are many facts pertaining to our embodiment in a human body, as Schechtman claims that this embodiment is inscribed in the institutions and practices that make up the person-space in which our lives are lived. “Our

psychological and social lives are thus infused everywhere with our biology” (149). Thus since PLV takes the cluster property model of person lives seriously it holds that these connections between different aspects of our lives as ordinarily lived constitute a defining feature of personhood and personal identity (149). As such, Schechtman notes two implications for understanding a person life. The first is that, unlike other analyses of personal identity, there is no single set of necessary and sufficient conditions required for a life to continue. Rather a life is a set of integrated and mutually-supporting functions. “When too many of these functions cease the overall integrity of the life can no longer be maintained and the life ends. So long as enough functions continue the life goes on” (156). The second implication is that person lives can be conceived to continue more or less robustly. “As functions are lost, the hold on life becomes more tenuous. An organism in which many life functions are compromised may be just barely alive, holding on by a thread” (157).

Schechtman uses the property cluster model to characterize the persistence and identification conditions for person lives. Recall the initial analysis we referenced in understanding personhood, that S is a person if and only if S is an entity who lives a person life and continues for the duration of a single person life. We determine persistence and personhood by determining whether the person life they are living continues and can be identified as the same person life. So in understanding whether the identity of persons can extend we can ask the separate question of whether person lives extend and infer an answer to the former question through the latter. “Since PLV takes the cluster property model of person lives seriously it holds that these connections between the different aspects of our lives constitute a defining feature of personhood and personal identity” (149).

Schechtman thus seeks to understand a person life through the property cluster concept that utilizes a constellation of different functions, biological, psychological, and social elements

that make up a person life. In analyzing the case of a human in a permanent vegetative state, she notes that a human in this state not only has a seriously compromised biological life, but an even more compromised person life. “Most of the functions that define such a life are lost when someone is in a vegetative state. Not only is there mitigation of biological functioning, there is a complete or almost complete loss of psychological functioning, and social relations are drastically compromised as the person in a vegetative state can be only the recipient of social attentions and not an active participant in interactions with others” (157).

Yet despite all of this, Schechtman still maintains that, according to PLV, a person who falls into a permanent vegetative state survives as a vegetative human person, though they are a person in marginal and unusual circumstances. A human that enters a permanent vegetative state may have at one point been considered a writer, for example. Their social interactions and their contributions to their person-space may have been such that both they considered themselves a writer, and also that the social infrastructure and people in their life supported the intuition that they are a writer. Upon entering the vegetative state, the human no longer has the psychological and social capacities to maintain this role. As such, we can say that, by Schechtman’s analysis, the human body *alone* has lost all relevant functions and capacities that would maintain its status as a writer. Yet, if we are to take Schechtman’s application of the property cluster as also working for the concept of a writer, then we could posit that the writer can still exist (albeit we have to grant relevant other properties for the property cluster of writer, and also allow that there are enough other relevant properties coexisting and interacting in relevant ways so as to posit the continued preponderance of this property cluster in the absence of the social and psychological properties forwarded by a human body). Sure, the writer may exist in a degenerate way, and that some of the important properties of the cluster are absent. But we can still posit the existence of enough of the properties necessary for the

cluster's survival, thus allowing that while the human body of the writer lacks the capacities necessary to actualize its identity as a writer, we can still ascribe the identity of writer irrespective.

For instance, we may think that a writer must possess the qualities 'frequently writes', 'understands how to construct prose', and 'can read' in order to be a writer. We may claim that these qualities are necessary qualities for the identity of writer. However, if we take the property cluster model we can still attribute the identity writer while certain properties of the cluster are not met. Thus, while 'can read' is a quality not met by John Milton, we can still see this individual as a writer for his production of *Paradise Lost*<sup>21</sup>. It is not too far of a stretch to posit that some of the qualities of being a writer are qualities not constituted by the individual we ascribe the title to. For instance, we may think that some of the properties of being a writer are 'affecting people with written material' or 'having work widely disseminated'. If we follow the property cluster model, then it is entirely possible that enough of the relevant properties for the cluster concept 'writer' are retained even in the absence of certain essential properties that were retained by the body. Thus, we can imagine someone being attributed the property of writer even though the individual's body does not contribute to the survival of the property cluster through the sustained cohesion and functioning of the relevant properties.

Likewise, there are things beyond the body that maintain the status of the vegetative human as a person. Schechtman claims that humans in permanent vegetative states are still persons, albeit severely limited ones. They are still persons because they are still included in a

---

<sup>21</sup> This example may be a little problematic. John Milton was blind when he worked together with his family to have his epic tale transcribed. The argument could be made that he fully could read, but simply lacked his perceptual faculty. However this does not defeat the push for the cluster approach: we could easily imagine an individual who utilizes aids and transcribing devices to pursue their identity as a writer without being literate.

person-space, as well as the fact that others will, as Schechtman says, continue the vegetative humans narrative. "Close friends and family who had relationships with the person usually continue their relationship with the patient" (105), be this through continuing to visit the vegetative human, overseeing their care, or decorating rooms in symbolically significant ways to reflect aspect of the persons life (105). In this way, Schechtman claims that people continue on the life narrative of the vegetative person and thus keep the personhood intact. If we allow that 'has a life narrative' is one of the properties that supports the property cluster, a claim that Schechtman supports, then we have good reason to think that this particular property of the cluster is maintained by things external to the body. The personal identity of the human that entered the vegetative state is thus maintained and enabled by externals, and thus in important ways we can see personal identity as transcending the body: there is the possibility that more than the organism is maintaining personal identity.

While the claim can be made that the persistence of relevant properties of writers and of persons is different, we can acknowledge an important conclusion. The point here is that it seems that persons and writers, by Schechtman's account, can have their persistence and identity conditions maintained irrespective of whether the human body cannot relevantly contribute and support the property cluster. Sure, with both concepts we can acknowledge that the human body is *still* contributing to the property cluster (say by maintaining the property of 'there is an alive organism'). However, it seems that the human body is not enough on its own to guarantee the survival of the writer or of the person. In this instance, we may still claim that the human body was the originating source of many of the properties necessary for the actualization of a person life, but also that many of these properties are beyond the body itself. We grant that the body is not enough to actualize all of these properties (and likewise all of the three components of a person life), and that in some cases the body may not be able to

actualize enough of these properties on its own to maintain a person life. Even though this is the case, we can still characterize them as living a person life.

In what follows a briefer discussion of individual analyses of personal identity and personhood will be undertaken. The goal of this endeavor is to show how other analyses can be reconciled with the thesis of extension. Part of this undertaking will analyze views that I believe are largely irreconcilable with extension. However, it should be noted here that my intention in presenting these analyses is not to assess them for their philosophical merit. I do not intend to analyze and compare the strengths of these views, but rather to assess their compatibility with the thesis of extension, as I have done with Schechtman's analysis.

## 2.2 Milojevic

Recalling the work done in the second chapter, Milojevic's main objective with her article is to provide a defense against the charges against HEC made by Miyazono. Thus her discussions on theories of personal identity and personhood are cursory. She still does provide a systematic investigation of these terms and the tenability of their extension. However, this discussion is not the focus of her larger project here.

Milojevic states that she will assume the psychological continuity view in answering questions about the persistence conditions of personal identity. Importantly she claims that the psychological continuity view is essentially functionalist, in that persons are identified by functional properties (16). In regards to Milojevic's take on the psychological analysis of personal identity and personhood, Milojevic understands persons to be special kinds of minds. Milojevic conceives of persons as changing entities whose identities are determined only by psychological continuity, independent of personal traits and specific parts of their bodies.

Prior to her discussion of persons, Milojevic makes the claim that there are hybrid integrated systems with mental properties. One example is of the Otto-notebook hybrid system, a system Miyazono differentiates from Otto. Miyazono had claimed that, by the analogy between Mark and the Mark-book system, the mental states of hybrid systems cannot be attributed to individuals in the system (beliefs are attributed to the system but not to Mark). Milojevic believes that the attribution of belief states to the Mark-book system is ill founded. She claims that Mark and the book—are not sufficiently integrated and since they do not satisfy the integration conditions she earlier established (11-13). Since the Mark-book system does not have any extended mental states, this implies Miyazono’s argument does not give us good reason to differentiate between Otto and the Otto-notebook system.

Granting the conditions of integration are rigorously met, Milojevic will make the claim that biological beings and hybrid systems can be identical with one another. Borrowing from psychological accounts of personal identity and personhood, we can claim biological Otto is identical to the Otto-notebook system in that they are both psychologically continuous with one another because of the chains of psychological connections between their mental states. “Otto’s occurrent beliefs produced all of the Otto-notebook system’s dispositional beliefs, wishes and desires of narrow Otto still guide the behavior of hybrid Otto, hybrid Otto remembers some of the experiences of narrow Otto, etc.” (18). Milojevic goes further with this assertion about extended personal identity. Not only is it that hybrid systems are identical to the original unextended cognitive subject, but also that in some cases we can even claim that the retention of personal identity is made possible by extension. In ways Otto’s personal identity is kept intact by the processes of hybridization, whereas without his notebook he would find his psychological capacities limited and he would find it harder to ensure the strength and retention of his identity as it is. In sum, Milojevic claims that, “if personal identity is sustained through

hybridization then mental states which are realized in both biological and artifactual parts of the system are attributable to the person that entered the process of hybridization and thus, constitute a part of the extended mind” (16). Likewise, if we take Milojevic at her word when she states that persons are simply special types of minds, and that we recognize minds on functionalist grounds, then there seems to be no reason to deny that persons do not extend on Milojevic’s account. “If persons are simply subjects of mental states...then if mental properties are widely realized, so is the subject of these properties” (22).

### 2.3. Olson and Williams

The animalist theory, as advanced by Olson, provides some very strong arguments to support the intuition that extension is not viable when considering personal identity and personhood<sup>22</sup>. Recall that the biological view is an analysis that asserts that all humans, and subsequently all person stages of such, are biological organisms. Personhood is not a substance category but a subsequent stage that is experienced by the human animal in its development. In accordance with the biological account, a human animal’s continued existence is based on the notion that it maintains biological continuity by means of its vital functions—these functions being the continued metabolism, teleology, and organizational complexity of the organism, among other things. One presupposition of Olson’s condensed synopsis of the biological criterion is that “anything that is an animal at one time will always be an animal” (1999 136), the implication here being that one cannot be an animal at t1 and an inorganic being at t2; if an animal ceases to maintain its vital functions it thus ceases to exist. Since an animal is contingent upon the vital functions that create life, an animal, when dead, is not a dead animal but the lifeless remains of an animal that no longer exists. Subsequently, the biological view of personal

---

<sup>22</sup> The animalist view discussed by Olson has already been discussed in chapter 3. As a result, our discussion of the account will be more cursory in nature here.

identity asserts that you are the animal (which may or may not be experiencing a person stage), and when this animal fails to exert vital functions it ceases to exist.

If the thesis of extension here is the assertion that personal identity and personhood can extend beyond the body, then it seems that the biological view is antithetical towards this possibility. Proponents of the biological view believe that we are and only will be biological animals. To posit that our personhood can be something beyond this animal is irreconcilable with most animalist views.

One thing Olson talks about in expressing his reticence about extension is the problem of too many thinkers. Essentially, if we posit that Otto is the hybrid integrated system, then he cannot possibly be an animal. An animal can never be anything but an organic entity, and thus allowing that Otto is his notebook leads to the conclusion that Otto cannot be that animal alone. This leads to the problem of which entity is thinking in these instances. Olson believes that we are forced to allow the existence of the thinking hybrid system and also of the organism that is Otto who is also thinking. If one decides that it is only the hybrid system that is thinking, then this means that organism-Otto is never alone thinking. This creates problems for anyone trying to reconcile the animalist view with extension.

Another theory that may significantly push back against this thesis of extension is an account we will loosely refer to as the bodily view. In his article "The Self and the Future" (1970), Bernard Williams articulates a concern for why we should reject philosophical accounts that prize psychological continuity over other possible criteria of personal identity. One way that psychological accounts have been advocated for is through appeal to problem-cases of body swapping. Essentially the claim is that body swapping procedures are described in such a way that drives the intuition that mental continuity and psychological features are the relevant

concern in personal identity. Because of the ways these descriptions are illustrated, “the philosophical arguments designed to show that bodily criterion was at least a necessary condition of personal identity would seem to be mistaken” (167).

Williams proposes that when the body swapping scenario is described differently, the intuition that is supported is that the bodily criterion is now emphasized in describing where our future concerns and perceived answers to the importance of the body for the identification of persons (167-174). Williams describes an iteration of the body-swapping experiment in such a way to illustrate that we may still rationally fear what happens to our bodies once our psychological character is separated from them. Williams describes a thought experiment involving body-swapping that supports the intuition that it is the body, not psychological continuity, that is what matters in our considerations of identity. Williams outlines a thought experiment where a scientist announces the following to you: you will be tortured, you will not remember any of the things you are now in a position to remember due to being put into an amnesiac state, and that you will have a different set of impressions of your past, quite different from the memories you have now. Since one may still rationally fear for this event in where the torture will happen, this supports the notion that we can still rationally fear and consider the future of our body when all psychological character has been eliminated. This, along with other arguments, is done to show that psychological criterion cannot be the only thing to be considered when asking questions about the future experiences of entities. “The principle that one’s fears can extend to future pain whatever psychological changes precede it seems positively straightforward” (180). The argument’s simplest iteration is that we can still rationally fear for the treatment of our body after psychological character is lost or significantly altered (so significantly altered that the psychological character before the change cannot be considered identical to whatever psychology remains/survives). This claim is advanced by the notion that

one's concern and fear for the future of their body is motivated by the notion that the fear is fear for oneself. The fear itself will be motivated by, "the thought of how it will be for me, the imaginative projection of myself as a participant in x" (175).

Recall our discussion of the role of procedural memory in discussions of identity. The analysis by Schechtman is carried out to show that the mind cannot be viewed as the brain alone. For our purposes, we used her analysis to show how procedural and non-semantic declarative memories can be important for understanding identity. Her analysis can support Williams' concerns here, in that we can think not only that the body remains an important locus of concern when evaluating our rational future concerns, but also that important aspects of identity are in the body itself and not just contained within psychological character. The above argument remains a strong case for considering the importance of the body in certain questions. Thus, we can conceptualize the bodily view as the view which states that we are our bodies and that our identity through time consists in the identity of our bodies<sup>23</sup>. This kind of analysis, like that of animalism, seems intuitively incompatible with the thesis of extension. If identity is maintained through the identity of bodies, then it seems that things beyond the skin-and-skull boundary cannot possibly ground identity. Likewise, if we are our bodies then it is impossible that personhood can be found in anything beyond the body.

Perhaps there is room within these bodily accounts for one to argue that the introduction of body-machine interfaces or silicon neural implants constitute part of the body,

---

<sup>23</sup> See also Ayer (1936) and Thompson (1997). While Williams provides strong reasons for considering the bodily account, he ultimately supports the claim that there more than just the bodily criterion is required in answering questions about personal identity. "A question of personal identity is evidently not answered merely by deciding the identity of a certain physical body. If I am asked whether the person in front of me is the same person as one uniquely present at place a at time t, I shall not necessarily be justified in answering 'yes' merely because I am justified in saying that this human body is the same as that present at a at t" (1956-1957, 229).

and thus can be included in the person and in identity. Likewise, perhaps the bodily view may be open to the body itself being extended in a radical way, perhaps thinking of the disembodied brain that connects with the body through radio waves as something of an extended hybrid system. However, if our ultimate goal is to use the intuition about body-machine interfaces and spatially separated organisms to jump to the conclusion that Otto's notebook is therefore part of him, then there is reason to anticipate heavy resistance from the bodily account. If the bodily account can only accommodate extension of the body into neural implants and prosthetics, then it seems that extension is allowed but only if the extension does not violate the skin-and-skull boundary. The allowance of this thesis does not prove anything radical that we wish to pursue with our investigations of personal identity and personhood.

Likewise, I doubt that theorists who defend either the animalist or bodily view would allow that an organism or body can have parts that are spatially separated from each other and yet still considered part of the same organism/body. Olson invites us to imagine a thought experiment where one's cerebrum is cut out of one's head and transplanted into another body. In answering whether the brainless animal is the animal, or if the transplanted cerebrum is the animal, Olson supports the notion that the brainless animal (if all its coordinated vital functions are not disrupted) has simply lost an organ. In terms of identity, the individual is the brainless organism, not the detached cerebrum. This is because, by the animalist account, we are animals and the animal that lost the organ could still be characterized as a living organism. Thus the original organism and the later organism with the lost cerebrum are identical, and for purposes of identification the person that was the original organism is the latter cerebrum-less organism, even if all mental functioning has ceased (and by this omission the organism is no longer a person). This view is supported by the notion that, "a detached cerebrum is not a living organism at all, any more than a freshly severed arm or detached kidney is an organism.

Although it is composed of living cells, it cannot coordinate its activities in the way that a living organism can” (Olson 107). By this omission, I think Olson would be inclined to say that, if the brain were removed but still connected with the body via a process of hybridized extension, then for Olson the animal would not be located in these two locations because the detached cerebrum cannot coordinate activities in the way a living organism can. It is along these same lines that Olson argues that, “the person who ends up with your cerebrum in the ‘transplant’ case could not be you, if you were an animal, for he is not the animal you are. He does not have your biological life” (107). Even though the bodily view may not make explicit reference to the persistence conditions of organisms, I do not believe that we have much reason to think that they differ from animalists in this respect.

In what follows is a discussion of a novel problem for theorists that attempt to reconcile extension with personal identity and personhood. This problem pulls from the work of Derek Parfit (1971).

### 3. Parfit

Derek Parfit provided some insight and complications for the concept of personal identity. Generally it is assumed that where there are questions about personal identity there must be an answer. It is believed that the nature of our identity is such as to guarantee that, in certain problem cases, questions about identity must have answers. Parfit notes that certain questions do presuppose an answer about personal identity, but believes that these questions can be freed from this presupposition (4).

Imagine a case of brain transplants. The functioning brain of Shelton is placed in a brainless body, and the resulting person has Shelton’s character and apparent memories. The claim by Parfit, and many philosophers and thinkers that subscribe to a psychological account of

personal identity, is that the resulting person is Shelton. Parfit proposes to expand upon this situation. Imagine that Shelton's brain is divided, and each half is housed in a new body. Both resulting people have Shelton's character and apparent memories of his life. Parfit proposes three possibilities for the answer to the question of identity. 1. Shelton does not survive; 2. Shelton survives as one of the two people; 3. Shelton survives as both.

The first possibility is implausible, considering that we initially allowed the full brain transplant earlier. Considering that individuals have survived with half their brain destroyed (5), why would it be that the individual does not survive? In regards to the second possibility, implausibility is also noted here. On what grounds would one survive as one but not as the other? And finally, the third can also be seen as implausible, but only under the caveat that one takes survive to imply identity. If survive implies identity, then the third possibility is implausible, for it is impossible for an individual to be two persons (5). If survive does not imply identity, then the third possibility is irrelevant to a problem about identity. After some discussion about this possibility of joint survival, of the existence of two people, Parfit moves on to an alternative possibility. He suggests giving up on the language of identity. Thus in response to the brain fission scenario, we can disassociate survival with identity and suggest that Shelton survives as both people without answering or presupposing the answer to the question of whether he is one of those people. "We can suggest that I survive as two different people without implying that I am these people" (8). Parfit claims that once we give up on the belief that there must be an answer to any questions about personal identity, that we will no longer encounter any difficulty in certain problem cases (in that we realize there is no answer to the problem case). Identity may be a one-to-one relation, but survival need not be. Parfit goes on further to qualify what this sort of concept of survival would entail and how the question of survival does not entail a question about identity.

The notion that questions of identity do not always have an answer allows for more clarity when talking about certain problem cases. Parfit discusses a malfunctioning teleportation device, that inadvertently results in the cloning of an individual. It is equally valid to eliminate the presupposition that we can provide an answer to this case instead of attempting to answer the questions of identity that pop up. However, even though some cases don't presuppose a question of identity, Parfit still believes that psychological continuity is a ground for speaking of identity when it is one-to-one (12). Psychological continuity itself need not be of a one-to-one relation, and as a result, "if psychological continuity took a branching form, no coherent set of judgements of identity could correspond to, and thus be used to imply, the branching form of this relation" (12). Thus, even if psychological continuity is neither logically, nor always in fact one-to-one, it can provide a criterion of identity (13). Parfit sketches the following analysis from the above: "x and y are the same person if they are psychologically continuous and there is no person who is contemporary with either and psychologically continuous with the other" (13).

The reading of personal identity in this manner raises a problem for views of extended personal identity. Personal identity is characterized by a one-to-one relation. It is numerical identity that allows us to assert claims about personal identity, and the violation of such that allows us to note that our ability to discern identity is problematized. It is the fact that a person cannot be two individuals at once that raises a difficulty for our ability to ascribe personal identity in certain cases. Imagine an entity that was able to 'split' into two autonomous and independent consciousnesses. Both resultant individuals would be functionally and qualitatively identical, however they would operate with individual consciousnesses and bodies. At a later point these two entities may remerge to form one entity, consolidating their experiences and recent memories within one entity. Parfit addresses these types of scenarios, and believes that while we can understand psychological connectedness and the criterion for identity, in these

cases we can only evaluate the survival of branching and merging psychological continuity: there is no answer for identity here.

Imagine another scenario. Take a similar sort of eldritch being, except instead of splitting into two individuals, the entity is able to split off individual mental functions (say their internal memory storage of the time they pushed a button). The result would be an individual who possesses all previous mental facilities, and an externalized process. We could imagine this kind of entity as having a machine encode their memories while simultaneously deleting them from the original organism. If personal identity is one-to-one and composed of psychological continuity, and at least half of an 'entity's' continuity is grounded in a different spatial location than the organism, then the notion is that the entity could be separated from this other half of its psychological continuity. In this way this entity, the 'resultant plus external storage', could feasibly be construed in the same way as the two resultant Shelton-entities from the brain fission case. Say this spatially divided entity has some form of interaction in which they cannot communicate with the other part that constituted continuity for the original subject. Does the resultant still retain the same personal identity when they do not have access to half of their psychological continuity? By Parfit's analysis, this may be a situation where the question of identity cannot be answered, insofar as there are multiple 'entities' that are psychologically continuous (or at the very least related) with each other. The only difference in this scenario is that instead of splitting into two entities, one entity is splitting its psychological continuity between a resulting descendant (though significantly different due to lacking continuity) and a form of storage to maintain the other half of psychological connectedness.

The main issue is that, if identity is a one-to-one relation, then how can we conceive of our identity as extended into other external artifacts? Does this not violate the numerical relation of identity, thus making the attribution of personal identity to individual and relevant

externals arbitrary? When talking about the identity of the resultant plus external storage, we cannot point to a place or spatially cohesive system that contains the individual. The individual is two things under a certain perspective, seemingly violating the numerical identity of the initial entity.

One initial solution is that, because there are no other entities that complicate a one-to-one relation, we can presuppose a question and answer to identity. Perhaps the answer is no, since there is no significant psychological continuity between the resultant half-entity and the ancestor. Or perhaps it is that identity is preserved, but only when the resultant is in communication with the other half of the psychological continuity. Parfit's understanding of the psychological criterion as the basis for personal identity is explicit in stating that the psychological criterion cannot be met if there are other entities psychologically continuous with the entity in question (13). There is no violation of Parfit's claim since there are not multiple entities in existence at one time. It is not the case that both the initial entity and the resultant plus storage system are both in existence at the same time. The claim ultimately would be that, if the glue and trust conditions are maintained in the proper way that emulates the functional description of atypical cognitive subjects, then the initial subject is identical with the resultant plus storage system, both by means of analyses of personal identity and by a functionalist reading.

I think the next possible solution to this problem is to deny that the numerical relation is violated. Instead of thinking of individual and externals as violating the numerical aspect of identity, why not instead think of personal identity as applicable to a complex entity? If we allow that the resultant entity and his external storage can be conceived of as a multifaceted entity, then given that there is no functional difference between their interactions and the ancestor's occurrent memory, we can infer that they are identical. In this case the numerical relation

between the complex entity and ancestor would be preserved, given the fact that the glue and trust conditions are met between the parts of the complex entity.

One possible objection to this sort of conclusion would be to deny that psychological continuity can be maintained between the resultant and the external storage since there are certain subjective psychological elements contained in only half of the resultant plus storage system. Because of this, it may be that the resultant plus storage cannot be numerically identical with the ancestor since psychological continuity is not fully established between the resultant and the external storage.

There are ways to push back against this claim. The first is to note that psychological continuity is kept between the ancestor and the resultant plus storage system even if the subjective psychological elements are only contained in half of the system. This is because psychological continuity may still be established between the resultant and the storage system. We may grant that only the resultant can be the author of subjective psychological elements, yet still note that these subjective psychological elements are informed by the psychological information kept within the storage unit. Remember that both the identity relation and psychological continuity are only maintained as long as the glue and trust conditions hold between the two: as such, the resultant may be the author of subjective psychological elements that are heavily informed by psychological elements contained within the storage. For example, say the resultant finds themselves looking at a painting. They may find themselves with a whole array of different emotions that are evoked upon witnessing it. Now imagine that the resultant is looking at this painting but is also part of the extended system, where the external storage has access to the knowledge that the painter is a family member of the resultant. It seems more than likely that, if we grant that the glue and trust conditions are sufficiently met, then the person observing the painting may feel a myriad of different emotions than they would have

without the knowledge of who the author of the painting was. Perhaps now they feel pride upon seeing the painting, or melancholy directed at how their relative has impacted the world. If we allow this example then even though certain subjective psychological elements can be actualized within the resultant, that psychological continuity and meeting of the glue and trust conditions can entail that the storage system is still important for the articulation and explanation of these subjective elements.

Likewise, there is more to psychological continuity than just the survival of subjective psychological elements, as our investigation on the role of procedural memory revealed. The intuition supporting the extension of Otto via his notebook does not rest upon the assumption that the notebook itself creates subjective psychological features. And yet the notebook can be considered to be an important part of the Otto-notebook system that itself actualizes the subjective psychological features. Even though the resultant and the storage system may split psychological continuity in such a way that certain cognitive and mental functions can only be feasibly created and sustained by one part of the system, this does not mean that the two parts are still not psychologically continuous with one another.

If we are to allow that this sort of dilemma about numerical identity of personal identity does not significantly jeopardize the possibility of extended persons and personal identity, then we may infer the following. First, we need not consider that individualism about persons is threatened by the possibility of extension. Persons remain individuals, even if composed by multiple 'parts'. We may even come to the conclusion that one's personal identity is grounded in the extension of memories via transactive memory systems between two or more people. This need not lead us to the conclusion that persons are not inherently individuals (that persons can merge in on each other or divide like cells), complicating our ability to distinguish between people.

The second implication is that, if we allow that personal identity and personhood is not threatened by the charge of confusing numerical identity, then the onus for proving that personhood and personal identity does not extend is upon individual analyses of personal identity and personhood. Specifically, theorists must point out why it is that their analysis will entail that the initial entity is not the same as the resultant plus storage system, or that Otto's personal identity and personhood do not extend into his notebook. This is not to say that this is an impossible endeavor. A theorist may push back against this sort of conclusion by positing a hybrid of a psychological account that also prizes the importance of a continuing body. This way the theorist can deny that extension entails that personal identity extends, while rejecting or allowing a modest hypothesis for HEC or the extended mind. However the onus is not on us to verify an account of personal identity or personhood that entails extension. Reasons have been given in this chapter to think that memories are relevant to personal identity and likewise that memories can be considered to extend. If one subscribes to an account that prizes the role of memory for personal identity, then reasoning will have to be given to show that extension does not affect personal identity.

#### 4. Conclusion

The previous chapter concerned itself with providing a somewhat rigorous independent analysis of the role of memory in theories of personal identity. If we grant its role in grounding identity in a substantive manner, and that there are good arguments for considering it as extended, then we have good reason to think that personhood and personal identity can extend beyond the body. With this work done, the purpose of this chapter has been to provide some analyses of various views on personal identity and personhood and their compatibility with extension. There is ample room for us to read extension into analyses of personal identity and personhood, and thus for us to begin to conceive more broadly the fact that we as entities

extend beyond our bodies. Yet, as we also saw, it is not the case that all analyses are compatible with extension.

A question emerges of how we ought to evaluate the capability to accommodate extension. Should we understand theories of personal identity and personhood that allow for extension as stronger, or as more viable? I do not necessarily think that this is the right course of action. The thesis of extension is supported by our acceptance of functionalism. This type of functionalism, when applied to theories of personal identity and personhood, allows for a semblance of multiple realizability. Thus the concepts that we apply our investigations to must allow a sense of multiple realizability in them. When it comes to considering mental events, there appears to be consensus in the literature that mental events can be multiply realized, even if there is still contention of whether we can absolutely assert this. However, when we consider what makes an organism, multiple realizability has its limits. The criterion for what an organism is admits of less open-endedness in realizing an organism in different forms. For what makes an organism the organism can be identified in its functions, yet it is not the case that the animalist would recognize the same organism if we were to replace certain architecture that still actualizes the initial function. While functions can be multiply realized, the animalist would deny that certain important architecture that ensures function F can be replaced by other architecture while maintaining the organismic integrity of the animal. Functional description is not enough to entail that something is part of an animal or not by the animalist account.

I believe thus that we should not consider accommodation with extension as a boon, or as a point in favour of the verification of the analyses, but rather as an implication in how our analyses entail that we view persons and personal identity. However, if we are to allow the veracity of the extended mind thesis then perhaps we have reason to think that the *proper* way to conceive of personal identity and personhood is as, under certain circumstances, extended. In

this case, we need not think that the animalist and bodily account, or the other accounts that may seem irreconcilable with extension, are wrong, but rather that they disagree with how lenient we ought to be in how we ascribe functionalist descriptions on the world.

What follows the end of this chapter is a brief discussion of what further claims this thesis might motivate. In particular, the coming conclusion will detail possible implications EST may have on our conception of extended entities of which persons are only parts and the way we consider collective mental states and how extended persons may influence this. Following this will be a brief discussion of one significant problem that faces us in our ability to conceive of extension in regards to personal identity and personhood. This problem focuses on a problem Parfit introduced to discussions of personal identity and personhood.

## Chapter 6. Conclusion

The previous chapter endeavored to analyze some accounts of personal identity and personhood through the lens of extension, working off of the groundwork laid in the previous chapters that attempted to open up the possibility of extension. If we grant the work of the previous chapter, then we have good reason for thinking that most versions of the psychological account are compatible with extension. Likewise, it was shown that there is room to argue for extension with accounts that do not explicitly endorse something of a psychological criterion. While there were views analyzed that seemed incompatible with extension, work was done to show that there is open space for us to analyze various conceptions of personal identity and personhood through the lens of extension validly.

The purpose of this thesis has been, in part, to analyze the foundation of EST, by looking at its foundational base, HEC, and arguing that claims of extension can be made validly. By analyzing HEC and arguing against those that attempt to refute it, I have provided a base for evaluating EST. In analyzing EST, the conclusion was reached that analyzing EST through the concepts of personal identity and personhood was an apt endeavor, one that may allow us to more successfully conceptualize the possibility of extension. This analysis of EST was furthered by investigating a few authors who oppose the possibility of extension. The conclusion was reached that their worries were not enough to completely trivialize the possibility of extension. In an attempt to reach a positive thesis, an analysis was undertaken of the role of memory in analyses of personal identity and personhood, primarily that of psychological accounts. If one allows the intuition about Otto's extension into his notebook, then there is room to argue that Otto's personal identity and personhood extends into the notebook. This intuition was strengthened by broadening the conceptual criterion of memory to include other forms of non-autobiographical memory, as well as referencing a few authors who argue for the extension of

autobiographical memory onto the environment. Following this, several investigations were carried out to analyze the possible accommodation of extension by other theories of personal identity and personhood. While it was shown that many accounts can be seen as compatible with the thesis of extension, two accounts were shown to be ultimately incongruent with extension.

This thesis is not arguing for the strong claim that personal identity and personhood ought to be conceived of as extended, or that the extension of persons and personal identity is confirmed. For, as I discussed briefly, the only way to verify extension in this context is to first argue for the veracity of a specific account. This thesis does not endeavor to provide this sort of investigation. Merely the claim that if one follows the logical progression of the preceding chapters, then we have good reason to think of various criteria and analyses as compatible with extension.

One necessary shortcoming of this thesis was the limited scope through which we analyzed theories of personal identity and personhood. Because this thesis endeavored to analyze HEC and EST in order to provide a solid basis to understand the extension of personal identity and personhood, less focus was provided on individual investigations into theories of personal identity and personhood.

It is because of this that this thesis should not be seen as a knock-down case for the extension of personal identity and personhood. This thesis only investigated a very small section of human considerations of personal identity and personhood. This thesis endeavored to focus on several contemporary considerations of EST, and to focus our investigation on popular contemporary accounts of personal identity and personhood within western philosophy. With this approach done, I believe that a decent sample of prominent theorists were addressed.

However, there is a large portion of the literature that went unaddressed. Not to mention, this thesis specifically focused on Western analyses of personal identity and as such neglected to talk about any Eastern accounts of personal identity and personhood. It is possible that, had this thesis taken the extra space to analyze more disparate accounts of personal identity and personhood, then we may have been left with a stronger or weaker impression about the tenability of extension. Regardless of this shortcoming, however, I believe that the goal of this thesis was met. The intention to analyze and investigate EST and to provide argumentation opening up the possibility of seeing personal identity and personhood as extended has been addressed here. Even if someone disagrees with the conclusions drawn here, this thesis provides a thorough analysis of various authors that talk of HEC, EST and of the extension of personal identity and personhood. This thesis may at least stand as good material to introduce and explain the concepts surrounding the concept of extension in these narrow domains.<sup>24</sup>

This concludes the thesis. While the work done in this thesis can be seen to have conceivably narrow implications for the way we conceive of philosophical personal identity and personhood, there are ways that we might expand these findings and apply them to different domains. For instance, if we grant the claim that social environments play an active role in sustaining cognitive activities and operations, and since minds can extend beyond the boundaries of skin-and-skull, then social institutions can play a constitutive role in the formation and exercise of mental activity. Mason Cash (2010) wishes to expand upon this claim to revise the individualistic way we perceive moral agency and responsibility, ultimately arguing that, “responsibility can be collective and distributed in ways that are quite unlike our typical

---

<sup>24</sup> With this all said and done, I do believe that this investigation into the extension of personal identity and personhood can have a larger effect beyond the literature discussed. If we are to take the conclusions of this thesis seriously, that the extension of personal identity and personhood is not only tenable but also possible in some cases, then perhaps we have room to further support arguments for collective entities and states.

accounts of individual responsibility” (659). Ultimately Cash believes that we can expand our position to view an autonomously ‘intelligent’ agent as one who, “critically engages the cognitive tools around them, one who selects, endorses, and uses effective cognitive tools, who replaces, refines, or augments less effective cognitive tools, and who selectively incorporates these social, relational technological, environmental, and bodily resources into their sense of who they are, what they know, what they want, and what they can do. Thus both cognition and autonomy are offloaded onto the social, technological, and normative environment” (661).

Ultimately Cash claims that we can best conceive of moral responsibility as both a cultivated competence and as an attribute that may not necessarily only be ascribed to individuals; moral responsibility can be ascribed to both entities larger than persons and to systems of which persons are only parts. Responsibility and agency thus seems to be something that not only can extend beyond persons, but is something that can be ascribed to entities larger than persons, like the abstract concept of a community’s normative practices and the people that maintain them (656). If we allow ourselves to seriously consider these claims, then there are a few predicaments we need to investigate. One is the possibility of having concepts that, while normally only attributed to individual persons, extend beyond the boundaries of persons to be best conceived of as social concepts. The second implication, closely related to the first, is of there being instantiations of collective mentality. Gallagher and Crisafi (2009) propose that cognition is extended into ‘mental institutions’ such as legal systems and museums. Institutions like these are both the products and producers of human cognitive activity. If we allow Gallagher and Crisafi’s claim that these institutions are appropriately linked with human mental activities like problem-solving, then they argue that these institutions can collaborate in hybrid cognitive processes that should still count as cognitive, even if not done in the head. If we accept the conclusions drawn from this thesis, then I think we have room to

seriously conceive of cognition and concepts like responsibility for actions as public, collective matters based upon shared normative practices and institutions—to which each individual action potentially contributes to and from which each action draws (Cash 664). The arguments driving the acceptance of HEC are functionalist by nature and, if we allow Gallagher and Crisafi's claim that similarity to internal cognitive mechanisms is not a good hallmark of cognitive processes (and thus reject the parity principle's argument from analogy<sup>25</sup>) then there seems even more possibility for the wider recognition of cognition and mentality<sup>26</sup>.

One further thesis, as discussed in depth by Huebner in *Macrocognition* (2014) is of the possibility of collective mental entities, even of the possibility of collective personhood. This is a difficult concept to parse. On the one hand, if we allow ourselves a functionalist conception of personhood, then there is room to argue for a positive reception of collective personhood. Petit (2007) claims that if an autonomous group faces a significant choice between morally valenced options, has the understanding and access to evidence required for judging the relative value of options, and has the control required to make judgments and act according to these judgments, then we can think of collective agents as responsible. Thus if we conceive of an analysis of personhood that can have all of its criteria collectively realized, then can we ascribe personhood in such a wide instance?

There are significant reasons for why we may push back against this possibility. Huebner notes that, "since there is nothing that it is like to be a collectivity (in the ordinary case),

---

<sup>25</sup> Specifically, Gallagher and Crisafi want to claim that we should reject the restrictions made by the parity principle. Instead of seeing human cognition as the benchmark upon which we understand cognition in the environment, we should expand our conception of cognition to be instantiated in non-human systems.

<sup>26</sup> There are problems for this approach. Recalling the work done at the end of chapter 2, there may be reason to hold to our characteristic description of cognition through analogy to human cognition, granting that, at this moment, human cognition is the benchmark by which we can extend our functionalist analogy.

extending cognitive capacities to collectivities seems to disrespect the fact that individuals are subjects of experience” (182). If we consider experience to be something intrinsic to being an individual, then it seems that collectives could never be individuals. Yet it seems like a misrepresentation to claim that a collective *is* an individual. We can grant that a collective group may not be the subject of experience in the same way that individual humans are, while still arguing for an analysis of personhood that does not presuppose experience as an essential criterion.

Regardless of the above, there are further difficulties with conceiving of collective personhood or personal identity. What happens to the personhood of those individuals that may be parts of this collective person? How might we understand the qualitative and quantitative identity of this type of collective person? Does our conception of numerical identity implicitly presuppose the dismissal of this type of person? And if it does, is it perhaps tenable to reject the notion that a person must only be numerically identical to themselves? These are questions that I believe that the work of this thesis has helped provide a starting point for in future discussion.

This thesis does not intend on providing a substantive answer to the question of the plausibility of collective mentality or of collective entities. However, if we grant the work done in this thesis, then I believe we have a proper foundation to discuss these possibilities. With all this said and done, however, the thesis will conclude noting the possibility of extension within some theories of personal identity and personhood, and that this possibility can aid in investigating further theses that may put pressure on individualism in personal identity and personhood.

That’s it.

## Works Cited

- Adams, F. and K. Aizawa. "The Bounds of Cognition." *Philosophical Psychology*, vol. 14, 2001. pp. 43-64.
- Addis, D. and L. Tippet. "Memory of Myself: Autobiographical Memory and Identity in Alzheimer's Disease." *Memory*, vol. 12, 2004. pp.56-74.
- Ayer, Alfred. *Language, Truth, and Logic*. London, Gollancz, 1936.
- Baddeley, Alan. *Working Memory*. Oxford, Clarendon Press, 1986.
- Baker, Lynne. "Persons and the Extended-Mind Thesis." *Zygon: Journal of Religion and Science*, vol. 44, 2009. pp. 642-648.
- Baker, Lynne. "Technology and the Future of Persons." *The Monist*, vol. 96, 2013. pp. 37-53.
- Barba, Gianfranco et al. "A longitudinal study of confabulation." *Cortex*, vol. 87, 2017. pp. 44-51.
- Benson et al. "Neural basis of confabulation." *Neurology*, vol. 46, 1996. pp. 1243-1249.
- Buller, Tom. "Neurotechnology, Invasiveness and the Extended Mind." *Neuroethics*, vol. 6, 2013. pp. 593-605.
- Carruthers, Peter. "HOP over FOR, HOT Theory." *Higher-Order Theories of Consciousness*, edited by Rocco Gennaro. Amsterdam, John Benjamins Publishing, 2004.
- Cash, Mason. "Extended Cognition, Personal Responsibility, and Relational Autonomy." *Phenom. Cognitive Science*, vol. 9, 2010. pp. 645-671.
- Chalmers, David. *Extended Cognition and Extended Consciousness*. Forthcoming.
- Clark, Andy and David Chalmers. "The Extended Mind." *Analysis*, vol. 58, 1998. pp. 7-19.
- Clark, Andy. "Intrinsic Content, Active Memory and the Extended Mind." *Analysis*, vol. 65, 2005. pp. 1-11.
- Clark, Andy. *Supersizing the Mind*. Oxford, Oxford University Press, 2011.
- Cohen, Neal. "Preserved Learning Capacity in Amnesia: Evidence for Multiple Memory Systems." *Squire and Butters*, vol. 1, 1984. pp. 83-103.
- Colheart, Max. "Confabulation and Conversation." *Cortex*, vol. 87, 2017. pp. 62-68.
- Cowan, Nelson. "The Magical Number 4 in Short-Term Memory: a Reconsideration of Mental Storage Capacity." *Behav. Brain Sci*, vol. 24, 2001. pp. 87-114.

- Dennett, Daniel. "The Self as a Center of Narrative Gravity." *Self and Consciousness: Multiple Perspectives*, edited by Kessel P. Cole and D. Johnson. Hillsdale, NJ, Erlbaum, 1992.
- Drayson, Zoe. "Extended Cognition and the Metaphysics of Mind." *Cognitive Systems Research*, vol. 11, 2010. pp. 367-377.
- Dritschel et al. "Autobiographical Fluency: A Method for the Study of Personal Memory." *Memory and Cognition*, vol. 20, 1992. pp. 133-140.
- Fitts, William and Louise Warren. *Tennessee Self-Concept Scale*. Los Angeles, Western Psychological Services, 1996.
- Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy*, vol. 68, 1971. pp. 5-20.
- Frost, Joshua. "The Extended Self, Functional Constancy, and Personal Identity." *Linguistic and Philosophical Investigations*, vol. 12, 2013. pp. 47-66.
- Gallagher, Shaun and Anthony, Crisafi. "Mental Institutions." *Topoi*, vol. 28, 2009. pp. 45-51.
- Galloti, Mattia and Bryce Huebner. "Collective Intentionality and Socially Extended Minds." *Philosophical Psychology*, vol. 30, 2017. pp. 252-268.
- Gertler, Brie. "Overextending the Mind?" *Arguing About the Mind*, edited by B. Gertler & L. Shapiro. New York, Routledge, 2007.
- Haris, C. B, P.G. Keil, J. Sutton and A.J. Barnier. "Collaborative Remembering: When can Remembering with Others be Beneficial?" *Proceedings of the 9th conference of the Australasian Society for Cognitive Science*, edited by W. Christensen, E Schier and J. Sutton, Macgaurie Centre for Cognitive Science, North Ryde, 2010.
- Heersmink, Richard. "Distributed Selves: Personal Identity and Extended Memory Systems." *Synthese*, vol. 194, 2017. pp. 3135-3151.
- Heersmink, Richard. "Extended Mind and Cognitive Enhancement: Moral Aspects of Cognitive Artifacts." *Phenom Cogn Sci*, vol. 16, 2015. pp. 17-32.
- Huebner, Bryce. *Macro cognition: A Theory of Distributed Minds and Collective Intentionality*. Oxford University Press, Oxford, 2014.
- Klein, Stanley and Shaun Nichols. "Memory and the Sense of Personal Identity." *Mind*, vol. 121, 2012. pp. 677-702.

- Kopelman et al. *The Autobiographical Memory Interview*. Windsor, UK, Thames Valley Test Company, 1990.
- Kuhn, Manford and Thomas McPartland. "An Empirical Investigation of Self-Attitudes." *American Sociological Review*, vol. 19, 1954. pp. 68-76.
- Lewis, David. "Survival and Identity." *Philosophical papers*, vol. 1, 1983. pp. 55-75.
- Lindeman, Hilde. "Holding One Another (Well, Wrongly, Clumsily) in a Time of Dementia." *Cognitive Disability and its Challenge to Moral Philosophy*, vol. 40, 2010. pp. 161-168.
- Locke, John. "Of Identity and Diversity." *Personal Identity*, Edited by J. Perry. Los Angeles, University of California Press, 2008.
- Lycan, William. "The Superiority of HOP to HOT." *Higher-Order Theories of Consciousness*, edited by John Benjamins, 2004.
- Menary, Richard. *The Extended Mind*. Cambridge, MIT Press, 2010.
- Miller, George. "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information." *Psychol. Rev.* vol. 63, 1956. pp. 343-352.
- Milojevic, Miljana. "Extended Mind, Functionalism and Personal Identity." *Synthese*, vol. 198, 2018. pp. 1-28
- Miyazono, Kengo. "Does Functionalism Entail Extended Mind?" *Synthese*, vol. 194, 2017. pp. 3523-3541.
- Nagel, Thomas. "Brain Bisection and the Unity of Consciousness." *Synthese*, vol. 22, 1971. pp. 396-413.
- Nozick, Robert. *Philosophical Explanations*. Cambridge, Harvard University Press, 1981.
- Olson, Eric. *The Human Animal*. Cambridge, Oxford University Press, 1997.
- Olson, Eric. "An Argument for Animalism." *Personal Identity*, edited by R. Martin and J. Barresi, Oxford, Blackwell, 2003.
- Olson, Eric. "The Extended Self." *Mind and Machines*, vol. 21, 2011. pp.481-495.
- Olson, Eric. "Was I Ever a Fetus?" *Philosophy and Phenomenological Research*, vol. 57, 1997. pp. 95-110.
- Parfit, Derek. "Personal Identity." *Philosophical Review*, vol. 80, 1971. pp. 3-27.
- Parfit, Derek. *Reasons and Persons*. Oxford, Clarendon Press, 1984.

- Persuh, M. et al. "Working Memory and Consciousness: The Current State of Play." *Frontiers in Human Neuroscience*, vol. 12, 2018. pp. 1-11.
- Petit, Philip. "Responsibility Incorporated." *Ethics*, vol. 117, 2007. pp. 171-201.
- Puccetti, Roland. "The case for mental duality: Evidence from Split-Brain Data and Other Considerations." *The Behavioural and Brain Sciences*, vol. 4, 1981. pp. 93-123.
- Rupert, Robert. "Challenges to the Hypothesis of Extended Cognition." *The Journal of Philosophy*, vol. 101, 2004. pp. 389-424.
- Schechtman, Marya. "The Brain/Body Problem." *Philosophical Psychology*, vol. 10, 1997. pp. 149-164.
- Schechtman, M. "Personhood and the Practical." *Theoretical Medicine and Bioethics*, vol. 31, 2010. pp. 271-283.
- Schechtman, Marya. *Staying Alive*. Oxford, Oxford University Press, 2014.
- Shoemaker, David. *Personal Identity and Ethics*. Toronto, Broadview Press, 2008.
- Shoemaker, Sydney. "Persons, Animals, and Identity." *Synthese*, vol. 162, 2008. pp. 313-324.
- Sprevak, Mark. Extended Cognition and Functionalism. *The Journal of Philosophy*, vol 106. 2009. pp. 503-527
- Theiner, Georg. "Transactive Memory Systems: A Mechanistic Analysis of Emergent Group Memory." *Review of Philosophy and Psychology*, vol. 4, 2013. pp. 65-89.
- Thomson, Judith. "People and their bodies." *Contemporary Debates in Metaphysics*, edited by T. Sider, J. Hawthorne & D. Zimmerman. Oxford, Blackwell Publishing, 2008.
- Tulving, Endel. *Elements of Episodic Memory*. New York, Oxford University Press, 1983.
- Tulving, Endel. "Episodic vs. Semantic Memory." *The MIT Encyclopedia of Cognitive Sciences*, edited by R. Wilson and F. Keil. Cambridge, MIT Press, 2010.
- Vold, Karina. "The Parity Argument for Extended Consciousness." *Journal of Consciousness Studies*, vol. 22, 2015. pp. 16-33.
- Wagner, Nils-Frederic and George Northoff. "Personal Identity and Brain Identity." *The Routledge Handbook of Neuroethics*, edited by Syd M. Johnson & Karen Rommelfanger. Routledge: New York, 2017.
- Walter, Sven. "Cognitive Extension: The Parity Argument, Functionalism, and the Mark of the Cognitive." *Synthese*, vol. 177, 2010. pp.285-300.

Wegner et. Al. "Cognitive interdependence in close relationships." *Compatible and Incompatible Relationships*, edited by W. Ickes. Springer-Verlag, New York, 1985.

Wheeler, Mike. "In defense of extended functionalism." *The Extended Mind*, edited by R.

Menary. Cambridge, MIT Press, 2010. pp. 245-270.

Williams, Bernard. "Personal Identity and Individuation." *Proceedings of the Aristotelian Society*, vol.57, 1956. pp. 229-252.

Williams, Bernard. "The Self and the Future." *The Philosophical Review*, vol. 79, 1970. pp. 161-180.

Wilson, Robert and Bartłomiej Lenart. "Extended Mind and Identity." *Handbook of Neuroethics*. Dordrecht, Springer, 2015. pp.423-439.