A Critical Enquiry into the Metaphysics for Mind Uploading

by

Paul Warby

A thesis submitted in fulfilment of the requirements for the degree

PhD in Philosophy

in the Department of Philosophy at the

UNIVERSITY OF PRETORIA

FACULTY OF HUMANITIES

SUPERVISOR: Prof Emma Ruttkamp-Bloem

March 2021

Acknowledgement

Firstly, I would like to thank and acknowledge my supervisor Prof. Ruttkamp-Bloem for her continued support and input throughout the process. I am in her debt.

Secondly, I would like to thank my wife, Faeeza Warby Mohamed, for her patience, insights, and encouragement.

Last, but not least, I would like to thank my family for their ongoing support and inspiration throughout the process.

Abstract

Mind uploading is a fascinating possibility that asks us to imagine a person being instantiated in a substrate other than their biological body, such as a person continuing in a robot/computer. The current thesis takes a philosophical stance and enquires as to what the nature of minds and persons would need to be in order for such a scenario to be feasible and presented multiple realizable physicalism and psychological continuity as the two primary categories of necessary and sufficient metaphysical conditions. The thesis explores the mind within the context of the mind-body problem and presents a novel version of physicalism (multiple realizable physicalism) as the central concept amongst disparate philosophy of mind stances that would allow for the mind upload. The current thesis also introduces various concepts (e.g., nomological boundaries) and emphasised others (e.g., category mistakes) to demonstrate the feasibility of mind uploading through the argument that the mind is likely multiple realizable and of a physical substrate (multiple realizable physicalism). The current thesis then, in relation to persons, explores the persistence problem and determines that a psychological (as opposed to biological) solution is the preferred stance that would allow for mind uploading. In essence, if a person is a mind (psychology) and this mind is said to continue, then it should not matter whether the mind continues in a biological body or in an alternative substrate (such as a computer). What matters is the continuity of the mind (the psychological solution).

The thesis integrates multiple philosophical stances using these two primary categories of multiple realizable physicalism and psychological continuity and presents various constraints that may emerge in relation to various preferences within these two primary categories. Furthermore, the current thesis argues that a possible solution to the mind–body problem and the persistence problem may be found in identifying both the mind and the person with the processes of the substrate. In this sense, both minds and persons are not the physical substrate that instantiates these phenomena (i.e., minds and persons are not the body) but rather these phenomena (minds and persons) are the processes instantiated. A person/mind *is* what the body does and not the body that performs these processes. This identification of processes could in principle allow for all forms of mind uploading. In addition the current thesis presents a novel view of the self (here meaning both the mind and the person) as relating to specific types of processes (efferent processes).

List of Contents

List of figures

List of tables

Chapter 1

1 A critical enquiry into the metaphysics for mind uploading

1.1 Introduction

This thesis aims to explore what the metaphysical necessary and sufficient conditions would need to be for mind uploading to be a successful project. Consider the science fiction scenario, person's mind occurs in a new body (e.g., a robot or a clone) other than in their original body (e.g., the biological body that the person is currently associated with). Would this be the same mind, the same person? If so, what would the nature of minds and persons need to be for such a question to be answered in the affirmative? This is the primary research focus of the current thesis and will be developed throughout this chapter (see, in particular, section 1.3.) and throughout the thesis.

The aim of this chapter is to explore the motivation for mind uploading, to define some of the preliminary terms, and to explore the academic context. The first and second sections address the cultural and historical context, where the idea of the mind uploading project (from here on termed the 'MUP') has emerged, as well as sketching some implications for humanity should it be achieved. The third section addresses the academic context of philosophy of mind and the need for the current thesis in this regard. The fourth section defines mind uploading in relation to other projects, such as whole brain emulation and the singularity. Section five introduces general metaphysical concepts and relates these concepts to the MUP and the current philosophical literature that addresses the MUP. Section six identifies other considerations relating to artificial intelligence (AI) and section eight concludes the chapter by offering a summary of the layout of the rest of the thesis.

1.2 The motivation for enquiry into mind uploading

We live in a world where technology impacts on many areas of life and one such area of increasing interest is at the interface between AI-based technologies on the one hand and our understanding of the mind, and the nature and limitations of humanity on the other. Projects that explore and develop aspects of technology related to the mind are increasing, both in number and in budget. The Human Brain Project (HBP, www.humanbrainproject.eu), which aims to integrate multiple fields of study and develop a simulation of the entire brain, has received a budget of €1 billion

from multiple sources, half of which is provided by the European Union. The Massachusetts Institute of Technology (MIT) has recently set aside US$1 billion to develop a new AI college (Knight, 2018), France is investing €1.5 billion of public funds (Rosemain & Rose, 2018) and China US$2.1 billion (Larson, 2018) to develop AI. The world also now has its first silicon-based citizen named Sophia; a robot made by Hanson Robotics (Waish, 2017). There is, therefore, an increasing need to better understand the nature and limitations of the mind in relation to technology.

One area of interest related to technology and philosophy of mind is that of mind uploading. Mind uploading can be defined as the migration of a specific person's mind (including all mental phenomena such as personal memory, desires, goals, and preferences) from one substrate to another; usually (although not limited to) from a biological brain to a non-biological computer (Chalmers, 2014).

1.3 Cultural context

The idea of mind uploading has taken root within modern culture partially through fictional works such as *The Matrix* (Wachowski and Wachowski, 1999) and *Altered Carbon* (Morgan, 2002), but also through movements such as trans-humanism (see for example, https://humanityplus.org/; Hansell, 2011). Other terms such as 'transformation technology', 'radical technology', 'liberation technology', and 'disruptive technology' may be associated with similar aspirations for mind uploading. The underlying idea is that humanity can be enhanced and progress through the use of technology, with mind uploading being one form of transcending the human condition of death as well as enhancing our capacities (Schneider, 2008). For some (e.g., Benedikter, Siepmann, & Reymann, 2017), the desire to transcend the inevitability of biological death may have had a traditional grounding in religious traditions (e.g., the Christian resurrection or Eastern religious views of reincarnation), but now emerges in the guise of science. If mind uploading is achieved, a form of immortality, or at the least some form of an extended life, would be feasible. The term 'life' in this context is seen to transcend purely biological notions (the traditional domain of life) and to incorporate a wider definition, such as artificial life or, in the case of mind uploading in particular, the potential extension of mental life through continued existence in alternative substrates after biological death.

The impact of mind uploading, if feasible, would be astronomical. Not only could (mental) life be extended, but the understanding of the mind that would result would, among other things, for instance, lead to treatment of psychiatric disorders and, therefore, alleviate much of human suffering. For example, knowledge of the mind sufficient for mind uploading would include, among

other things, knowledge of mood and perception. Therefore, psychiatric disorders that include perceptual disturbances (e.g., schizophrenia) and mood abnormalities (e.g., bipolar mood disorder) would be better understood and direct system re-engineering or reprogramming could potentially be initiated within the partial upload (cyborg) or completed upload. There would also be financial and social implications. Would there be a need for life cover, or would the same economic system be in place and be redirected to "re-life cover" (to cover the costs of uploads)? Would the world become overpopulated, or would uploading allow for greater interstellar exploration? Would there be animosity between those who choose to remain biological without upload and the cyborg/uploaders? Would the uploaders have the same rights as the biologicals? The current thesis does not aim to explore the social (e.g., ethics, politics, and economics) implications of mind uploading and focuses rather on the necessary and sufficient metaphysical conditions for successful mind uploading.

1.4 The academic contexts

Should mind uploading be achieved in the future (and agreed to have been successful by all parties), it would have developed significant insight, if not resolved, into some of the most persistent problems of contemporary metaphysics; in this thesis, the focus is primarily on the mind–body problem and the persistence of identity problem. Alternatively, should mind uploading be metaphysically impossible, much financial expenditure and human effort could be redirected to simulation rather than emulation projects (defined in section 1.4).

The mind–body problem enquires how the mind (the mental) relates to the body (the biological/physical) when they *seem* to be so different. For McGinn (1989, 2007), as a physicalist, the core of the mind–body problem is how mental phenomena (the mind) can arise from brain processes (from a body). How can physical properties, such as neurons firing in sequence, lead to the experience of falling in love, or planning a thesis? For others (e.g., substance dualists such as Taliaferro, 2018), the problem relates to how and why the mental (as a non-physical substance) and the physical can and do interact. The point here is that if mind uploading were achieved, the mind of the individual (including these mental phenomena such as the experience of being in love) would be replicated through technology and, therefore, be better understood, which may then lead to dissolving the centuries old mind–body problem. The mind–body problem as it relates to the MUP is the subject of chapter 2.

The persistence problem of personal identity attempts to address the issue of how a person can persist (be the same person) over time, despite the fact that persons change over time. In what manner is the younger self the same as the older self? Can we survive biological death? What

conditions/features/attributes need to continue over time in order for the claim of the same personal identity to be acceptable? Which is more important, the continuity of physical matter (the body) or continuity of the minds (the psychology)? If mind uploading were achieved, we would gain significant insight, as well as the need to address new ethical issues that relate to persons who exist in non-biological (or alternative biological) forms, which may shed light on these questions. The persistence problem, as it relates to the MUP, is the subject of chapter 3.

Both academic proponents (e.g., Chalmers, 2010) and critics (e.g., Pigliucci, 2014) of mind uploading have acknowledged that there is "relatively little" (ibid, p.119) academic philosophical writing on the topic and that much of the discussion currently occurs within non-academic settings.[1] Furthermore, limited direct development of the subject of metaphysics of mind uploading is available (e.g., Corabi, & Schneider, 2012; Wiley, 2014) with some of the current literature (e.g., Eliasmith, 2013; Bostrom and Sandberg, 2008; Sandberg, 2013) assuming, or succinctly identifying, metaphysical claims within the context of more practical concerns, such as ethical constraints and the technological competencies needed. Therefore, a thesis that critically considers aspects of the metaphysics relevant to mind uploading is an appropriate and fruitful area of academic enquiry with potential for novel contribution to the field.

In order for mind uploading to be feasible, it needs to rest on an appropriate metaphysics of minds and persons. Thus, the problem focus of this thesis is the claim that mind uploading rests on two metaphysical premises or claims which function as necessary and sufficient conditions (see section 1.5.3 regarding necessary and sufficient conditions) for successful mind uploading; namely a multiple realizable physicalist account of mind–body relations and an account of transferable personal identity (psychological continuation preferred over biological continuation). They are sufficient conditions because they guarantee a successful mind upload from a metaphysical perspective (implying no need for any other metaphysical conditions), and they are necessary because wherever there can be said to be a successful mind upload, these two conditions will be satisfied.

In terms of the first condition, a multiple realizable physicalist account of mind would entail a physically constituted mind (physicalist) that would not be inextricably bound to a particular physical substrate (multiple realizability). This is a necessary and sufficient condition because first, if the mind is not physical it would fall beyond the purview of any future science and, therefore, the mind uploading artifact (e.g., the robot or clone) would not be able to constitute the same

1 It is beyond the scope of the current thesis to postulate why there has been relatively little philosophical work on mind uploading. It may be because philosophers simply have alternative interests (e.g., the ethics of AI), that mind uploading has not yet gained enough traction in society to elicit the interest of the philosophers, or any such hypothesis.

mind. Second, if the same mind is bound to a particular physical substrate (e.g., the current biological body), then, any artifact, even if it could constitute a mind, would not be the same mind. This does not constrain the realizability to either a computational (e.g., silicon based) or a biological (carbon based) substrate and, therefore, does not commit to a particular kind of physical substrate. 'Substrate' is a term used within the mind uploading literature (e.g., Astakhov, 2008; Corabi, & Schneider, 2012; Walker, 2014) and may be taken to be synonymous with substance for our purposes (see section 1.5.2 and chapter 2). Whereas 'substance' may refer to fundamental substances (e.g., physical), the term 'substrate' refers to the particular substance wherein a phenomenon is instantiated. For example, within a physicalist stance, the body is the current substrate of the mind/person. In a mind upload, the idea is that an alternative substrate could instantiate the same mind/person, such as the idea of uploading your mind to a robot, in which case, the robot would be the substrate that instantiates your mind (mind upload). It is not yet determined precisely how the physical monist substrate results in a mind and, therefore, it should not commit to a particular substrate type at the outset of the enquiry. The MUP presents three broadly defined substrate options: (A), an artificial/non-biological substrate; (B), a biological substrate; or (C), the project is metaphysically contra-indicated as being a physical substrate. These will be referred to throughout the thesis as the 'ABC options'. Options A and B are both physical (however defined), whereas C asserts that the mind is not physically dependent and, therefore, beyond the purview of artifact development.

The MUP is concerned with the creation of an artifact that could instantiate a mind/person in an alternative substrate. Artifact creation relates to the nature of substrates (the whole and components of the artifact), as well as to what the substrate does (the functions of the artifact). In relation to the MUP, the enquiry is, therefore, into what kind of substrate is needed and what does it need to do for it to be said to instantiate the same mind/person across substrates.

In relation to the second condition of persistence of identity, or transferable personal identity, this thesis aligns with the psychological solution to the persistence problem, which asserts that for persistence of personal identity, what matters is mind (psychology) rather than matter (the particular biological substrate).. This is a necessary and sufficient condition because for a successful mind uploading project the 'person' needs to continue across substrates and this cannot happen if the 'person; is material; thus, it is the mind (psychology) that needs to be uploaded (continue) into an artifact. This thesis is, therefore, a critical investigation as to whether these two premises would provide the necessary and sufficient metaphysical conditions for successful mind-uploading.

1.5 Defining mind uploading (whole brain emulation)

As stated earlier, mind uploading can be defined as the migration of a specific person's mind from one substrate to another; usually (although not exclusively) from a biological brain to a non-biological computer (Chalmers, 2014). The current thesis defines the mind at the outset in the broadest possible terms. It is taken to include both conscious and non-conscious entities and interactions. The concept of the mind at issue here, thus, includes but is not limited to, aspects such as memory, emotions, intentionality, propositional attitudes, cognitive capacity, and so on.

The terms 'mind uploading' and 'whole brain emulation' are sometimes used interchangeably within the literature (e.g., Chalmers, 2014; Sandberg, 2013; Bostrom, 2014), although variation in meaning may occur and, therefore, a clarification is needed. Bostrom (2014, p. 40) identifies three types of emulation, namely: 1) high-fidelity emulation, which emulates the "full set of knowledge, skills, capacities, and values" (ibid.) of a given mind; 2) distorted emulation, which has a "non-human" element but is able to perform the "same intellectual labour" (ibid.); and 3) generic emulation, which is akin to an infant that may develop adult-like functions. Within the current thesis, 'emulation' is associated with high-fidelity emulation (irrespective of specific technology such as an artificial or biological artifact), in that the aim of mind uploading is to transfer or replicate a specific person's mind. Lastly, note that emulation is distinguished here from simulation, in that an emulated system replicates all the necessary and sufficient conditions of a system, whereas a simulation simply mimics aspects of a system. For example, a weather simulator does not get wet, whereas an emulator would. The current thesis has, therefore, distanced itself from using the terminology of simulations because the goal of the MUP is to replicate/upload the person, in whatever medium or method, then emulation is a preferred nomenclature. Others may prefer alternative terminology. For example, Wiley (2014) refers to virtual brains (the computational pattern) as 'simulation', whereas elsewhere he terms mind uploading as 'emulation'. The distinction for Wiley appears to be, although not explicitly stated, whether the computerised mind is interacting in a virtual environment (where simulation is used) or in the natural environment (where emulation is used).

Wiley (2014) presents four possible procedures for mind-uploading. First, he considers temporal factors, such as that of gradual and scan-copy uploading. In gradual uploading, parts of the brain are gradually replaced until such as a time as the whole brain and its functions are retained in the technological artifact. Second, he considers that of scan-and-replicate (scan-copy) procedures. Within this process, there is an immediate/instantaneous replication of the mind from the

original biological substrate to a technological artifact. Temporal concepts such as 'immediate'/'instantaneous' are placed within the human time frame and are used as relatively fluid concepts related to the experience of the proposed uploaded mind. Applicable to both the gradual and the scan-copy scenarios is the question whether these uploads are destructive or non-destructive, i.e., whether the original substrate is destroyed or remains intact. The current thesis presents these upload scenarios as the MUP matrix (Figure 1-1).

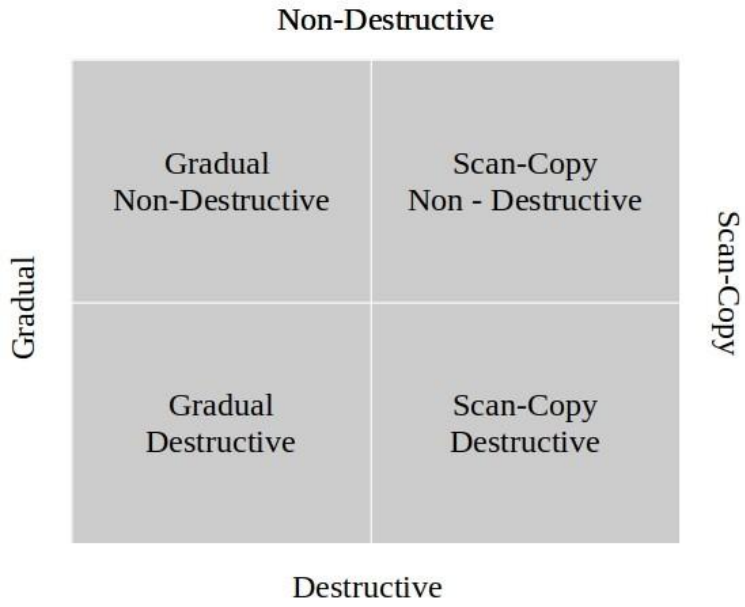

Figure 1-1: The MUP matrix

For the current thesis, the precise procedure of mind uploading is not the focus, however, the MUP matrix scenarios are used at times to illustrate or argue for a metaphysical possibility. For example, Chalmers (2014) asserts that a gradual upload, where the individual is conscious throughout the process, will serve to validate the process through conscious self-reporting during the upload process. The current thesis acknowledges that there may be multiple procedures to achieve mind uploading and, thus, rather orientates the enquiry into the metaphysical questions related to the end objective of a completed migration.

Another term commonly associated with, although distinct from, mind uploading is the 'singularity' (Vinge, 1993; Kurzweil, 2005). The singularity may be defined as the event in history when technological advances lead to a human-initiated artifact, or artifacts, that supersede human intelligence. The underlying philosophical assumptions are that intelligence (and, by implication, all types of mind) are a form of computation. When AI supersedes our biological capacities, it would be a feasible, absent defeaters, option (if the philosophical assumptions are reliable) to merge our minds with computers (mind upload) and so move humanity beyond our biological constraints

through increased intelligence and possible digital immortality. The singularity is, therefore, a concept that incorporates multiple facets and ideas related to technology, including mind uploading, but covers wider notions such as kinds of minds and the impact of technology on other spheres of life (e.g., economic, social, and political implications of AI), as already alluded to in section 1.2 above.

Mind uploading is a speculative technology and it is acknowledged that our current technology is insufficient to attempt the process. The current thesis does not aim to directly address the technological aspects of mind uploading, nor to predict any aspect of how or when this could occur, rather the aim is to enquire as to the metaphysical necessary and sufficient conditions under which mind uploading may be feasible, irrespective of technological progress.

1.6 A clarification of terms

Metaphysics can be defined as the "philosophical speculation beyond the current or seemingly possible limits of science, and the development of more or less abstract systems intended to explain ... the phenomena of mind and matter, and the place of man in the universe" (Gregory, 1998, p. 481). Metaphysics is, therefore, a significantly broad and complex philosophical subject and the current thesis orientates itself within the philosophy of mind, in the context of reflection on mind as a metaphysical problem in relation to the MUP. That metaphysics is, among other things, a speculative project enquiring into the limits of science, and the phenomena of mind and matter indicates that metaphysics is an appropriate field to explore the project of mind uploading. The two primary metaphysical questions that are raised in relation to the MUP are: What is the nature of a mind in the context of the mind-body problem (chapter 2) and what is the nature of a person in the context of personal identity as it relates to the persistence problem (chapter 3)? These two questions need to be answered in such a way as to allow for the feasibility of the MUP.

Furthermore, a distinction can be made between metaphysical, nomological, and physical necessity (Welshon, 2011). The distinction is based on which 'laws' are broken leading to a declaration of falsehood. The metaphysically necessary are bound only by the laws of logic and, therefore, may refer to possible worlds scenarios. The nomologically and physically necessary are bound by the laws of science and physics respectively (here categorised together) and relate more directly to the world (universe) we inhabit. Throughout the thesis (unless specified), the physical and scientific are categorised together as nomological and refer to the theories of science (which, in turn, refer to research and theories of the physical), and the various domains of sciences are termed nomological domains (e.g. physics, psychology, biology, and so on).

If the distinction between the nomological and metaphysical is accepted, it may be seen as an asymmetric relation, in that the nomological (that which is possible in our world) are, by implication, part of the metaphysical necessity (possible in all worlds), although the reverse may not be true. Therefore, nomological assertions may be used to infer metaphysical subject matter. Furthermore, current nomological concerns may give assistance when further speculation happens (such as that which occurs in metaphysics), through providing an empirical trajectory of verifiable theories. This empirical trajectory offers direction based on the success of other theories and inferring probable areas for future investigations and likely metaphysical conclusions. The current thesis uses this notion of the empirical trajectory throughout as the preferred epistemology, to establish what a metaphysic of the mind/person is.

In relation to science or possible technological developments, metaphysical enquiry is not primarily concerned with the particulars of how a specific scientific project could be ascertained (Heil, 2012), but rather with what the plausible necessary and sufficient conditions that may potentially guide a scientific project may be. The current thesis explores the possible integration of an individual mind with technology and, therefore, disciplines such as neuroscience, computer science, and psychology may be drawn from and reconciled with the ordinary experience of persons. Within the context of mind uploading questions over the nature of the mind and its potential replication in alternative substrates, as well as the result of a particular replication of personal identity, are, therefore, relevant metaphysical questions. The question of how a person functions is a scientific question, whereas the question of what a person actually is and in what sense this can be understood, is a philosophical (metaphysical) one. Therefore, the sciences play a role within metaphysics but do not dictate the process or agenda.

Furthermore, the current thesis introduces the distinction between what is a metaphysical problem and what is an engineering problem. Consider a man in the fifteenth century imagining a machine that could perform mathematical calculations at speeds and magnitudes that no person at the time could perform. This man may look at the machines of his day (e.g., a water pump) and declare that the calculating machine would be impossible. However, it would be a mistake to consider this a metaphysical impossibility (as modern machines are able to perform these feats) but should rather be considered an engineering problem based on the complexity of the machine and the limits of the science of his day. The issue of complexity is itself complex (Gell-Mann, 1994), but here is meant to illustrate the difference between a metaphysical problem and an engineering one. If a project is extremely complex it may be beyond the purview of artifact production, but this engineering problem is distinguished from the metaphysical problems that are the concern of the

current thesis. Presented below are some metaphysical concepts and how they potentially relate to the subject of mind uploading.

1.6.1 Metaphysical stances in philosophy of mind (substance and personal identity emphases)

The current thesis explores the two primary metaphysical problems of what a mind is (the mind–body problem) and relates this problem to questions concerning substrates (substances) and what substrates do; and the second metaphysical problem of what a person is and relates this problem to personal identity and persistence of identity through time.

In the context of the first problem of what the nature of the mind is, in relation to substrates (with the focus in the current thesis of the nature of the mind), the two dominant metaphysical stances within philosophy of mind are, broadly speaking, that of physicalism (here including naturalism) and dualism. These stances and their relation to the MUP are the subject of chapter 2 and are, therefore, only briefly discussed here. For the MUP, given that its focus is on artifact creation, the enquiry as to what the nature of a substrate is constituted by and what the substrate does, are the primary two concerns. For dualism, the mind is constituted by some ethereal (non-physical) substance. A non-physical substance lies definitively beyond the purview of technology and, therefore, no amount of human industry would be able to sustain a mind, because the entity that would be attempted to be produced would lie outside of technological endeavours. The type of dualism that is likely to nullify the plausibility of mind uploading is that of substance dualism (further developed in section 2.4.1), established most famously in modern times in the works of Descartes (1637/1968), and more recently revisited by others (e.g., Popper, 1953, 1955, 1983; Lowe, 2006; Loose, Menuge, & Moreland, 2018; Taliaferro, 2018).

According to Dennett (1978a, p. 252) "... it is widely granted these days that dualism is not a serious view to contend with, but rather a cliff over which to push one's opponents ...", and this tactic has been used by critics of mind uploading (Cappuccio, 2017; Benedikter *et al*., 2017). However, as Kim (2011) asserts, it would be a mistake to reject dualism *a priori* without serious discussion and, therefore, it is addressed, albeit briefly, within the current thesis.

On the other side, there are monist views of the mind. The most important one in our times, and also the dominant one[2], is physicalism. For the purposes of mind uploading, this is an important view, because the objective of migrating a mind requires that the nature of mind is a kind of a

2 In a survey of philosophers relating to the mind, "physicalism 56.5%; non-physicalism 27.1%; other 16.4%" (Bourget & Chalmers, 2013, p.15).

substance that technology can engineer. Because of this, physicalism is emphasised within the current thesis. Physicalism, first known as materialism until the 20th century, has a long history in Western philosophy from early Greek atomists such as Epicurus and Democritus to Hobbes *Leviathan* (1651/1973), writing in the relative time of Descartes.

According to Stoljar (2017) the term 'physicalism' was first introduced in the 1930s through the work of Carnap and Neurath and extends beyond matter (such as the precursor materialism suggests) *per se* to include objects of anti-matter, gravity, as well as including all facets of science that focuses on the dynamics and elements of the objective world. Hempel (1949/1980), asserted that the primary debate in philosophy of mind lies between that of physics (the natural sciences) and psychology (sciences of mind and culture), with science emphasising causal roles and theories. There are different variants of physicalism (to be explored in chapter 2), from reductive views based on some kind of identity between mind and body, through to non-reductive physicalism including property dualism and emergent views. As suggested already, what is significant to mind uploading may not be the exact nature of how the mind is constituted in the brain (an engineering problem), but rather the substance and nature of the mind (the metaphysical aspects of it) that can be technologically replicated.

For proponents of the MUP, physicalism is assumed. For example, Bostrom and Sandberg's, (2008), report on whole brain emulation (which includes, but is not limited, to high-fidelity emulation) briefly consider philosophical assumptions that are presented as primarily physicalism (that everything supervenes on the physical). These are simply stated without further discussion. Wiley (2014) states:

> "I will interpret the ideas of this book as primarily monist, i.e., claiming that reality is of a single (physical) substance" (p. 114).

More recently, Andrade (2018) affirms that a materialist (physicalist) conception of the mind is needed for uploading, although he does not develop this. Chalmers (2014) also affirms the mind as a natural phenomenon which is physical (see also section 2.7.1).

Critics of the MUP (e.g., Hauskeller, 2012; Corabi and Schneider, 2012; Pigliucci, 2014; Cappuccio, 2017) also emphasise that physicalism (reaffirming the dominance of the view in philosophy of mind) as necessary for the MUP but deny the feasibility on grounds that minds/persons are non-multiple realizable (e.g., Hauskeller, 2012; Cappuccio, 2017) or that that psychological continuity is not sufficient for persons (the subject of chapter 3). These criticisms, therefore, of multiple realizability (see chapter 2, in particular section 2.2) and psychological continuity (see

chapter 3), bolster the current thesis' claim that both multiple realizable physicalism and psychological continuity are necessary and sufficient conditions for the MUP.

Multiple realizability is usually associated with the physicalist stance of functionalism (although the current thesis will demonstrate that multiple realizability may be expanded to include various physicalist stances). Functionalism (e.g., Putnam 1960, 1965; Levin, 2018) is the view that defines mental states and events (mental phenomena) as functions that are descriptions of the causal roles enacted within the system of the mind. Importantly, a function is not limited to the substrate it is currently instantiated in (if another substrate can perform the same causal function this is sufficient) and, therefore, functions have 'multiple realizability'. This implies that an animal unlike us (e.g., an octopus), or an as yet unknown alien species, could have similar mental states to human mental states (e.g., pain), despite being constituted of different substrates. Moreover, if two systems are sufficiently functionally isomorphic (have all the same relevant functions), then the two systems can be said to have the same mental states and behaviours, the same mind. Both proponents (e.g., Andrade, 2018; Bostrom and Sandberg, 2008; Chalmers; 2014) and critics (e.g., Hauskeller, 2012; Pigliucci, 2014; Cappuccio, 2017) accept that functionalism, or some variation thereof, appears fundamental to the question of mind uploading. Functionalism and objections to functionalism are addressed further in section 2.6 of the following chapter.

A variation of functionalism is the representational theory of mind developed, among others, by Fodor (1974; 1975), and which is associated with the computational theory of mind, where symbols (representations) are manipulated to achieve outcomes (see section 2.6.2). The computational theory of mind (Rescorla, 2017) was developed in relation to advances in computational science asserting that the mind is itself a computational system. The central concept here is to see the mind in terms of a computational process, where information is processed through symbolic representation (syntax).

In relation to the representational variation of functionalism, it is noted that within the mind, uploading literature proponents (e.g., Andrade, 2018; Bostrom and Sandberg, 2008; Chalmers 2014; Wiley, 2014; Cerullo, 2015) and critics (e.g., Pigliucci, 2014; Hauskeller, 2012; Cappuccio, 2017; Benedikter *et al*., 2017) orientate mind uploading within a computational framework. Some (e.g., Pigliucci, 2014; Wiley, 2014; Cappuccio, 2017) briefly mention the possibility of a biological artifact that could, in principle, warrant discussion within mind uploading but this does not serve as the basis for their discussion. The current thesis, however, has not limited the MUP to non-biological artifacts (i.e., the MUP may include the biological artifact of the ABC options) and, therefore, presents these as engineering problems rather than metaphysical ones. If it is acknowledged that a mind may be instantiated in a biological artifact (e.g., Pigliucci, 2014), then

mind replication is a question of engineering (what material components are needed) rather than metaphysics.

Another contemporary physicalist view is that of property dualism, where there is one substance, the physical, but this substance is viewed to have two distinct sets of properties, one which is physical and one which is mental. Two pathways for the mental properties to exist are presented within the current thesis; namely that of emergence (e.g., Morgan, 1927; Emmeche, Køppe & Stjernfelt, 1997; Stephan, 1999) and that of panpsychism (e.g., Chalmers, 2014, 2015). On the emergence view (further developed in sections 1.5.6 and 2.7.4), mental properties are taken to emerge from lower level standard physical properties (e.g., energy, mass) and the lower level interactions among such properties. In panpsychism (see section 2.71), mental properties are taken to exist alongside standard physical properties. In both pathways, mental properties are taken to exist as something beyond standard physical properties but are still physical properties, nonetheless. In this thesis, this notion of 'beyond standard physical properties' is termed as 'further physical properties'.

In terms of the second metaphysical problem considered in the thesis, the nature of persons and the persistence of their personal identity are to be addressed within the current thesis in chapter 3, and we are confronted with two primary solutions, that of the biological solution (where persons are identified with their biological bodies – e.g., Williams, 1973; Olson, 2007) and the psychological solution (where persons are identified with their minds – e.g., Parfit, 1984; Lewis, 1987). In essence, to identify a person, supporters of the biological solution encourage us to follow the body and advocates of the psychological solution encourage us to follow the mind (see chapter 3). There is, therefore, a natural affinity between the psychological solution, where the mind is the necessary and sufficient condition for personal identity, and the MUP (which aims to upload a mind). Furthermore, within the thought experiments produced by proponents of the psychological solution (e.g., Locke, 1689; Parfit, 1984 – see section 3.4.) the nature of the particular substrate does not matter, as within these thought experiments the psychology (identity of person) exists in another substrate and is still deemed to indicate the same person (e.g., minds swapping bodies, minds in machines, and so on).

Critics (e.g., Hauskeller, 2012; Pigliucci, 2014; Cappuccio, 2017) and proponents (e.g., Chalmers, 2014; Walker, 2014; Wiley, 2014; Cerullo, 2015) of the MUP agree that continuity of personal identity (the metaphysics of persons) is essential to uploading (i.e., if no continuity is established, then in what sense could the mind continue across substrates?). Critics acknowledge that minds may be instantiated in alternative substrates, but then go on to deny that this would result in the same person persisting. For example, Pigliucci (2014) mentions the thought experiment of

Captain Kirk from the Star Trek franchise who teleports either to or from the star ship Enterprise. For Pigliucci, the teleported Kirk is a replica Kirk and not the same person, because the replica Kirk is constituted by a different body to that of the pre-teleported Kirk. In following the body (the biological solution), replica Kirk may feel the same, think the same, look the same, but because persons are their bodies this is not Kirk. In contrast, proponents of the MUP would assert that it is the same Kirk as there is psychological continuity (e.g., see Parfit's 1984 teleporter thought experiment in section 3.4.). Cappuccio (2017) echoes this denial of the MUP, based on the biological solution where personal identity is "embedded in the original brain-organism system" (p. 9), as does Hauskeller (2012), who asserts that in terms of mind as "situated in a body" (p. 7) the "only selves we have ever encountered are situated, embodied selves" (p. 7). Therefore, critics of the MUP base their denial of personal continuity in the notion that personal identity is synonymous with the continuity of the body (the biological solution), whereas proponents of the MUP base their affirmation of the MUP in the notion that personal continuity is a matter of continuity of the mind (the psychological solution).

In the next section, the current chapter aims to orientate the reader to philosophical concepts that have been adopted and adapted by the current thesis to these two problems of the mind–body problem and the persistence problem.

1.6.2 Substance, entities, and their interactions

"The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term" (Sellars, 1963, p.35). A distinction is made, therefore, in that there are "things" (entities) and there is "how they hang together" (interactions/processes). This approach does not presuppose what is accepted as entities, nor what is defined as processes, and allows for the possibility of the mental being an interaction (process), an entity, or both. Traditionally, minds were thought to be separate from the entities of the physical (e.g., Descartes, 1637/1968) and, as alluded to in the previous section, this raised the question of substance (substrate) in relation to the mind.

Kim (2011) defines 'substance' either as something where properties (and, by implication, interactions) are instantiated, or as a capacity for an entity to exist independently. Both of these definitions can be applied to the mind–body problem. Is the body the substance/substrate that the mind emerges from (or is there some particular mind substance), and can the mind exist independently of its physical substrate (the body)? Can the mind exist only in its current biological substrate, or can it exist in a new substrate? The current thesis does not attempt to resolve the mind–

body problem, but rather explores these fundamental theories and arguments as they relate to mind uploading, as well as contribute possible avenues of exploration and understanding.

In discussing how the mind relates to the physical, a caution is raised in terms of what Ryle termed the category mistake (1949). This kind of mistake occurs when an entity/interaction is incorrectly attributed to one logical category when it belongs to another. One example Ryle presents is of a person being shown around the university of Oxford or Cambridge. The person is shown the libraries, various departments, and the administration buildings but proceeds to state they have not yet seen the university. The mistake is in assuming that there is an as yet unseen category of the university that stands above the parts (e.g., the various departments). In terms of the mind–body problem, a category mistake may be made in conflating the substrate (an entity) with what the substrate does (the processes). For the persistence problem, a similar category mistake may be made in assuming that personal identity should be attributed to the substrate (the body which is an entity) rather than to what the substrate does (the processes).

1.6.3 Necessary and sufficient conditions (NSC)

The notion of NSC come up both in discussions of the mind–body problem and of the persistence problem and is generally understood as the sum of all criteria/conditions that would be needed for a certain entity/interaction to exist or for an event to occur. Copi and Cohen (1980) relate NSC to causation (further discussed in section 1.5.8.), which serves as an area of exploration for substrates and what substrates do within chapter 2. In relation to causation, a necessary condition is that which, if absent, would negate the possibility of the event. For example, in order for combustion to occur, a necessary condition would be the presence of oxygen. However, oxygen as a condition is insufficient condition for combustion, in that there are many instances where oxygen is present and combustion does not occur. Thus, a sufficient condition is that under which the event must or is guaranteed to occur (whether logically or contingently). In the example above, the sufficient conditions would be that a sufficient temperature is present in conjunction with other conditions, such as combustible substrate, oxygen, and so on. Therefore, there may be multiple NSC needed to be present for a particular event to occur. A necessary condition may, therefore, not be sufficient and sufficiency may refer to a collective of conditions that culminate in the overarching sufficient condition. For that reason, the current thesis uses the notion of necessary and sufficient conditions to explore what the metaphysical conditions for the success of the MUP are, without predetermining whether a singular or collective of necessary conditions is to be asserted.

In relation to identity (section 1.5.4. below), there are also NSC postulated by various theories for persistence of identity to occur. One method to establish NSC is to hypothetically remove, or add, conditions and monitor the outcome. Within the philosophical tradition (as well as the natural sciences), there are multiple forms of thought experiments (Brown, 1995) relating to this kind of hypothesising. For example, if it is asserted that a person requires a perfectly functioning body (here setting aside the question of what would be needed for a body to be perfectly functioning) to be a self/person, a thought experiment could be set out to explore what the result to self/person would be if the body was damaged (e.g., a limb removed). As there are many persons who retain their sense of personal identity despite having a damaged body, we can infer that the person who lost a limb would still retain their personal identity. As with the causal conditions above, identity conditions need not refer to a singular condition and may refer to collective conditions. Consider a bicycle. The bicycle has pedals, wheels, steering, and a frame. The wheel is not a bicycle, and the bicycle is not a frame, yet these are necessary for the bicycle phenomenon to exist. Alternative, a particular condition may be both necessary and sufficient. For example, a necessary condition of a water molecule is that it is constituted by two hydrogen molecules connected to an oxygen molecule (H_2O). Metaphysical conditions related to personal identity will be addressed within chapter 3.

The term 'necessary' may also be used in contrast to 'contingent'. A necessary truth is one where it would be logically contradictory to hold the opposite and is closely related to the notion of an analytic truth.[3] The statement that all bachelors are men is a necessary truth, in that the word bachelor has as a necessary facet the implication of being a man. A contingent truth is one neither impossible nor necessary but rather is dependent (contingent) on aspects of events as they occur, as is more closely related to synthetic truths. A contingent claim within a physicalist philosophy of mind might, therefore, assert that a certain neural state relates to a certain mental state, but that the particular neural state is not necessary for the same mental state. In contrast, a necessary claim within a physicalist philosophy of mind would state that a particular neural state always relates to a particular mental state by logical necessity. In essence, necessary and sufficient conditions are those conditions under which the phenomenon may exist and without those conditions would not.

3 The analytic/synthetic distinction is similar to the notion of necessary/contingent truths and relates back to the work of Kant (1781/1959). An analytic proposition is understood as true, based on evidence such as the concept of the predicate being included in the concept of the subject (e.g., "all bachelors are men"). A synthetic truth relies on other forms of evidence. For example, "the bachelor is tall" requires other criteria, such as the measurement of his height contrasted to that of the measurement of the other bachelors' height.

1.6.4 Identity

For one entity/interaction to be said to be identical to another is to assert that they are the same. For the current thesis, the topic of identity may be discussed within both the mind–body problem and the persistence problem. In the context of the mind–body problem, the subject of identity commonly relates to whether the mind may be identified with (identity claim) neural or behavioural activity (i.e., what the substrate does internally and externally). Whereas within the context of the persistence problem, identity relates to whether the person can be identified with the body (the biological solution) or the mind (psychological solutions). A common general distinction within the context of the notion of identity (irrespective of whether it relates to the mind–body problem or the persistence problem) is between a numeric identity and a qualitative identity, according to which the qualitative share some properties/processes, whereas the numeric shares all properties/processes (Noonan & Curtis, 2018). For example, two people can own the same model of car, sharing the same car design, manufacture process, metal properties of the chassis, and so on (qualitative identity), yet each car is numerically different (they can be separated in space and time and can, therefore, be assigned numeric designators). This can be seen as the classic view of identity and, if accepted, it could be asked whether the mind is best understood as a numeric or qualitative identity.

Suppose there was a software program running on one computer and the same program running on another. In one sense they are identical, in that they perform the same functions and respond in exactly the same way (qualitative identity). Yet they are numerically different, in that the one computer occupies an alternative place in space-time (different space-time continuity of the substrate). If the agenda is to perform certain computational functions, the numeric identity can be seen as trivial, whereas the qualitative identity is meaningful. On the other hand, if the agenda is to establish ownership, the qualitative identity is trivial and the numeric identity is meaningful. Suppose, furthermore, that the one program becomes corrupted and no longer functions. The substrate space-time continuity of the numeric identity is still intact but the qualitative identity is not. Which of these conditions now matter? Substrate space-time continuity appears trivial. Should we not rather view the qualitative identity and not the numeric identity as the NSC for a computer program? The current thesis argues that minds/persons are similar to computer programs, in that qualitative identity (as opposed to numeric identity) is the preferred approach.

Furthermore, the classic view of identity as either being numerical or qualitative has been contested. For example, by Geach's (1973) notion of relative identity (see also Griffin, 1974), where an absolute (strict) identity is denied as being possible (see also, Evans, 1978; Tye, 2000).

Gibbard (1975) presents his argument of 'contingent identity' using an example of a piece of clay that is a statue (similar to Aristotle's argument regarding form and matter[4]). The clay could be remade into an alternative statue (or just be a lump of clay), and so the identity of the clay statue cannot be emphasised as the substrate (the clay) but is rather to be seen as how the clay hangs together. However, the statue cannot exist without a substrate (the statue as a non-physical object), and so the statue may be seen as contingent on a) having a substrate and b) how the substrate hangs together (see section 3.5. on the Ship of Theseus for a similar argument). Furthermore, the statue could be reconstructed from metal or some other substrate and, therefore, settling the question of identity should not consider an emphasis on substrates (although a substrate is necessary) but on how substrates hang together. The implication for the MUP if this contingent identity is retained, would then be that the person/mind would relate more readily to how the substrate hangs together than the particulate substrate.

Furthermore, the current thesis presents a novel form of identity categories by distinguishing between partial and absolute identity, with partial identity implying that continuity of some conditions are necessary and sufficient, whereas absolute identity implies that all conditions are necessary and sufficient for continuation of identity (developed in section 3.2.). Partial identity is, therefore, akin to both qualitative identity and the alternative identity theories (e.g., relative /contingent identity) because only some (partial) conditions are necessary and sufficient for identity attribution to be asserted.

One way to formulate necessary and sufficient identity (and more relevant to the objective of a technological engineered artifact) would then be:

> x and y are identical when any necessary and sufficient property or process that is instantiated in x also may be instantiated in y.

The definition allows for both the numeric and qualitative identities to be asserted (e.g., the biological solution may assert that a particular body is the necessary and sufficient condition between x and y) and allows a spectrum of conditions from absolute (i.e., all possible conditions) to partial (i.e., any number of necessary and sufficient conditions).

In contrast to the notion of partial identity, the notion of absolute identity may be related to the indiscernibility of identicals (sometimes termed Leibniz's law). This law states that:

---

4 This has been termed hylomorphism and developed further in modern philosophy of mind, for example, see Jaworski (2011; 2016).

> If x and y are identical, then every/any property that belongs to x also belongs to y (Forrest, 2020). If it is asserted that two entities/interactions are identical, a path to refuting this claim would be to identify a property that is held by x but not by y (or vice versa). In short, if you can tell them apart (discernible) they are not the same (identical).

One method to sort through what are useful ways to identify phenomena is through so-called 'sortals' (see e.g., Grandy, 2016) according to which a phenomenon can either be sorted by substance (substance sortals) or phases (phase sortals). A substance sortal is that which is essential to the phenomenon for its existence (e.g., a gold ring cannot be the same entity if it were made of silver; it would cease to be a gold ring and be something else), whereas the phase sortal is true for some period of time (e.g., a person may be a child in one phase and an adult in another yet be the same person). Therefore, for the substance sortal the identity of the subject in question emphasises what it is made of, its substance, and the essential properties of the entity, whereas the phase sortals focus on what the subject does, its processes within a context, and as to what identifies it. Therefore, substance sortals relate to the substrate, whereas phase sortals relate to what the substrate does.

The notion of identity is further developed in chapter 3, as it relates to the persistence problem of personal identity. How is it that a person can change over time and yet be the same (identity preserved)? The problem is relevant to both the biological solution (e.g., the biological substrate changes molecules and atoms are constantly changing) and the psychological solution (e.g., our experiences of today influence the beliefs and desires of tomorrow and, thus, changes). A person is never the same person over time and never returns to a previous state precisely, and this change underpins the persistence of identity problem (how we can persist yet be different over time). Therefore, there may be some conditions that are necessary and sufficient conditions needed in order to retain identity (partial identity).[5]

In relation to mind uploading literature, the nature of personal identity is associated with the biological solution for critics and the psychological solution for proponents. For supporters of the biological solution (e.g., Corabi and Schneider, 2012; Pigliucci, 2014; Benedikter, *et al*., 2017;

5 Lewis (1987) presents a similar NSC, termed 'tensed identity', where what makes personal identity important is retention of "at least some significant class of properties" (ibid., p. 64). The current thesis expands this NSC to include both properties and processes as potential conditions for identity claims. See chapter 3.

Cappuccio, 2017), numeric identity (see section 3.5) is an essential feature of persistence of identity, as the person is identified with the continuity of the numeric substrate (i.e., the biological body that has continuity in space-time and is numerically distinguishable from other substrates). Because the MUP has persons being instantiated in alternative substrates, critics of the MUP would deny that the person is continuing because there is an alternative substrate. In contrast to the biological solution, advocates of the psychological solution (e.g., Chalmers, 2014) may allow for both numeric and qualitative identity. For example, if the numeric psychological solution (e.g., Lewis, 1987) is maintained, then so long as the psychology continues in one (numeric) substrate at a time (i.e., allowing for transfer of instantiation from one substrate to the next), the MUP is feasible (such as in destructive upload). Alternatively, the qualitative psychological solution allows for multiple substrates to instantiate the same (qualitatively) person (see sections 3.5).

The above brief overview of identity indicates that certain options are potentially open to mind uploading. The relevance of numeric, qualitative, or relative/contingent/loose identity may be explored under the larger rubric of absolute and partial identity. Once partial identity has been established as the preferred stance for the MUP, the question of which conditions are necessary and sufficient to identify a mind (chapter 2) and to identify and person (chapter 3) should be settled, as should be the question whether these are the same conditions.

1.6.5 Types and tokens

Another metaphysical distinction (Wetzel, 2018) that has been commonly used in the mind uploading literature (e.g., Walker, 2014; Wiley, 2014; Cappuccio 2017) is the type–token distinction. Generally, the type is a more abstract and unified concept, whereas the token is an instance of a concrete particular.

To illustrate suppose the following example:

‘Cat cat cat.’

In the above sentence, the word “cat” can be seen as a type where three tokens are instantiated. Furthermore, the letter “a” is a type with three token instances (as is the letter “t” and the letter “c”). The type and token distinction may have application at various levels, in that the token of a letter can be included in the type of a word, which, in turn, could have further types. Furthermore, different languages may use different words to refer to the same semantic meaning of “cat” and, therefore, the type may become increasingly abstract. A metaphysical question may be raised as to whether a type may exist independently of its token (the ancient universal/particular debate)

and could be associated with a form of platonic idealism[6] (Wiley, 2014). Alternatively, the reality of the abstract type can be questioned and only the tokens taken to be metaphysically real, where the types are simply ways of categorising (it is assumed for some purpose) tokens – this relates to nominalism[7].

Regarding the mind–body problem, the type–token distinction often relates to whether a mental phenomenon exists (e.g., pain as a type) and how the phenomenon can be identified with a particular neural correlate (the physical tokens of the substrate parts). For the type identity theorist (Smart, 1959; Place, 1956, see section 2.5.2.1.), each mental phenomenon type is the same phenomenon (state/event) and the same (identical) as the type of a neural activity (in the literature the common example used is of pain being identified with C-fibre stimulation). On the other hand, the token identity theorists (e.g. Davidson, 2001), who focus on relations of supervenience rather than identity, assert that physical tokens reflect the more abstract mental types (e.g., pain) but are not identical to them and, therefore, there need not be a direct one-to-one correlation between mental type and physical token. For Kim (2011), the type identity theorist can be seen as reductionist and the token identity theorist as a non-reductionist in relation to the mind (the mental) and the body (the physical). The type–token distinction here is, therefore, orientated to how tokens of the substrates' activities (e.g., neural activity) relate to the type of a mental phenomenon. The type–token distinction as it relates to the mind–body problem is further addressed in sections 2.5.2.1.

Regarding personal identity and the type–token distinction, the notion of token applies to the substrate as a whole (the body) and the metaphysical nature of persons applies to the notion of type. Those who hold to the biological solution emphasise that personal identity relates to the body (e.g., Williams, 1973) and they assert that the person type is bound to the particular substrate (the

---

6 Platonic Idealism (Ferrari & Griffith, 2000) relates to Plato's notion of forms, where an ideal exemplar of sorts stands beyond the physical world that participates in the exemplar. The metaphor of Plato's cave further reflects this. In this metaphor, we are asked to imagine a light projecting from the mouth of the cave, objects and interactions midway in the cave, prisoners who are bound and can only face the back of the cave, and the shadows that are cast on the back of the cave. The physical world is thought to be the reflected shadows that reveal the form of the real objects yet are not themselves the objects.

7 Nominalism (Rodriguez-Pereyra, 2019) is the philosophical stance that only the individual/particular/concrete (that which can exist independently of the mind) are real and, therefore, denies universals or abstract objects as real. A universal is an entity that may be instantiated by different objects (and, therefore, relates to the above discussions of types), whereas an abstract object is that which does not occupy either time or space (associated with Platonic Idealism) and is causally inert. Therefore, nominalism only accepts language descriptions that refer to individual/particular/concrete entities, or that are reducible to interactions of these entities, as being metaphysically real.

type = token), whereas those who hold to the psychological solution (e.g., Parfit, 1984) emphasise that the person type is not the token (type≠token). If the person type is not bound to the token, it allows for the possibility that the person type may be instantiated across substrates (similar to Gibbard's statue identity occurring in alternative substrates). The type–token distinction is reflected in the mind uploading literature by both critics (e.g., Hauskeller, 2012) and proponents (e.g., Walker, 2014; Wiley, 2014). The type–token distinction, as it relates to personal identity, is further developed within the thesis in section 3.6.

Furthermore, types may refer to descriptions of tokens sharing certain qualities and relate to the concept of similarity and distinction related to identity (e.g., a car may be more an abstract type where all cars are included as tokens or a particular model may also be a more restricted type where only car tokens of the same model are included).

The less strict the abstraction of type becomes, the more tokens are allowed within the sortal category of the type. The type can then be seen as an elastic sortal term that includes less or more qualities under the category used for identification. The current thesis introduces the type sortal spectrum that includes, on the one end, the type 'physical phenomenon type' (here assuming physical monism) and, on the other end, is a particular phenomenon that is absolutely unique, the 'absolute unique type'. The absolute unique type has no tokens that share any similar quality so as to be sorted into a type. To accomplish this, imagine a particular object that shares no similar qualities (i.e., properties or processes) with any other physical object in the universe. This absolute unique type occurs only once and is, therefore, an absolute numeric type. In considering persons/minds, the question then can be applied as to whether the particular person/mind type is such an absolute unique type or whether there is the possibility that alternative tokens may instantiate that particular person/mind (this is further explored in section 3.6). The type sortal spectrum is a descriptive spectrum and makes no assertions, or denial, of types as causal (i.e., if a type were to be described as an abstract object or universal, the spectrum may retain the description of the type and remain uncommitted to whether the universal/abstract object type is causal). Whereas as the type sortal spectrum is purely descriptive, the notion of a nomological boundary to be discussed below is associated with explanations within various nomological domains.

1.6.6 Nomological boundaries, reduction, and emergence

As stated earlier (section 1.5), nomological domains may be distinguished from the metaphysical yet may provide direction to the metaphysical project through empirical trajectories. The current thesis introduces the concept of nomological boundaries as a way to discuss both the necessary and

sufficient domains (including levels and stance on reduction) for the success of the MUP. The nomological boundary may be presented as delineating the domains that are needed for the replication of a system. In artifact design, necessary and sufficient knowledge and skills may be limited to a particular domain, or domains. Consider a carpenter who may proceed with his artifacts without any knowledge of all the nomological domains relevant to carpentry. For example, the carpenter may be unaware of the atomic theory that explains the solidity of the wood, the geological theory of how the metal was formed to produce the nails, or the evolutionary theory that explains how the wood evolved. What this example demonstrates is that for artifact creation the artisan needs no infinite knowledge of all relevant domains but may limit knowledge to certain domains, and this is what this thesis refers to as the nomological boundary (the boundary of what is necessary and sufficient knowledge for artifact design and creation).

Nomological domains may be seen as relating to each other vertically or horizontally. A vertical relation implies that the levels relate to higher or lower domains. For example, atomic theory is at a lower level then molecular theory, which is at a lower level than biological theory. Horizontal relations refer to domains that fall alongside each other to varying degrees. For example, mechanical and electrical engineering are similar domains on the same level and are closely related, whereas biology is on the same nomological level (that of the everyday world) but is less closely related. For the MUP, the question may be asked as to what are the necessary and sufficient domains for the replication of a mind/person, i.e., where or what are the nomological boundaries for 'you'?

In discussing levels of nomological boundary, we may consider the concept of emergence. Emergence can be broadly defined as the moment when a system exhibits novel features/properties when it exceeds a certain level (tier, order) of complexity. In essence, entities and their interactions at one level can result in new entities 'over and above' the ones these entities emerge from. To add to Sellars' notion of "things" and how they interact, this thesis states that, at times, things interact to form new things (e.g., atoms interact to form molecules). Emergence, in general, can be seen on a spectrum from strong (irreducible) to weak (reducible) (Stephan, 1999) emergence[8]. Regarding philosophy of mind, strong emergence allows for the possibility of a mind/person to be constituted by, emerge from, standard physical properties. These mental properties, although they emerge from

8 Stephan makes a distinction between weak emergence and two forms of strong emergence, namely, synchronic emergence and diachronic emergence. Synchronic emergence is related to the relation between the systems' properties and its micro-structure, irrespective of time, and this is thought to be irreducible (strong emergence). Diachronic emergence relates to the predictability of a novel property over time, that could not be foreseen before the first instantiation and is, therefore, also irreducible (strong emergence).

standard physical properties and processes, are not reducible to them. On the other hand, a view of weak emergence maintains that, although certain properties of the mental may emerge, they are nevertheless reducible, i.e., the claim is that higher level phenomena (e.g., psychology) can be reduced to lower level phenomena (e.g., neural activity).

Aside from strong and weak emergence (both views acknowledge that a new property/entity emerges), a notion of practical emergence may be put forward. Such a view accepts that levels may be a helpful way to explain and describe phenomena at various levels, while denying the ontological reality of the emergent property/entity. For example, Kim (2010) asserts that distinguishing between levels may be useful in terms of a nomological domain, but that there is the possibility of "cutting across levels" (p.64), which would negate that a "comprehensive ontology of levels" (p. 64) is necessary.

The current thesis acknowledges that many philosophical stances do not engage neatly with discussion on levels and reduction. For example, a panpsychist such as Chalmers (2015) may assert that mental properties are non-reducible, not because they emerge from an alternative level but because they are distinct properties at the same level (a horizontal dimension). Analogously, just as electrons are not reducible to neutrons (they are distinct physical entities with distinct properties), so it is thought by some that mental properties cannot be reduced to standard physical properties. Eliminativists (e.g., Churchland, 1981; Churchland, 1986), discussed further in section 2.5.2.2, deny that there is anything to reduce (mental theories should be eliminated and not reduced). For the MUP, what matters is that the mind/person are replicable irrespective how one wishes to view the issue of reduction.

The notion of a nomological boundary offers a way to circumvent many of these difficulties. Consider emergent reducibility (the distinction between weak and strong emergence) in relation to the nomological boundary for carpentry. Assume a particular table design requires that the table be constructed from material that has the property of solidity. Although solidity is a reducible property, what would the implications be if it were a non-reducible property, such as a strong emergent property? The carpenter need not concern himself with these questions of reducibility as the nomological property of solidity (based within a certain level of scientific domain) is necessary and sufficient for the table artifact replication. The carpenter need have no knowledge whatsoever of atomic theory or quantum mechanics (all of which are scientifically validated knowledge that relate to the property of solidity). The nomological boundary (the domain where the solidity of wood is a property) acts as a criterion for necessary and sufficient knowledge irrespective of questions of reducibility or the form of emergence (weak or strong) that is preferred. Furthermore,

whether one accepts non-reducibility through the assertion of a horizontal property (see above regarding panpsychist property dualism) or deny that anything needs to be reduced, such as the eliminativists, also does not influence the application of the nomological boundary. In each of these instances, the philosopher is free to determine what they deem to be the necessary and sufficient nomological domain (e.g., the neural activity of the eliminativists or mental properties and their interactions of the panpsychist property dualist) in both the context of the mind – body problem and the context of personal identity. And they are free to apply these nomological domains as the necessary and sufficient conditions for determining the nature of the mind/person. The notion of a nomological boundary, therefore, offers a neutral way of discussing the nature of the mind and person without *a priori* committing to a particular philosophical stance.

1.6.7 Manifest and scientific images

Within the philosophy of mind, there is often an emphasis on the disparity between how things appear (e.g., subjective experience such as qualia, or persons as entities) and how things are best understood through the sciences (see sections 2.3). This relates to a further distinction made by Sellars (1963) between the manifest and the scientific image. The notion of the manifest image, broadly speaking, relates to entities as they appear to us (e.g., tables, persons, family, money, countries). Within the everyday life of a person, these images are uncontested as part of our world. The notion of the scientific image relates to entities and interactions that are based on scientific theories and models (e.g., quarks, electrons, molecules). For Sellars (ibid.), the distinction reflects modes of enquiry where the scientific image is the primary candidate to contribute to our understanding of metaphysical reality.

Within philosophy, various stances align to this distinction to varying degrees. For example, closer to the notion of the manifest image are philosophies such as 'naive realism' (where reality is defined, absent pathology such as hallucinations, as it appears to us), whereas closer to the notion of the scientific image are philosophies such as scientific realism (where reality is more closely related to scientific theories and models, see e.g. Chakravartty, 2017).[9]

9 Both of these philosophies can be contrasted with anti-realist perspectives. For example, empiricism and instrumentalism (e.g., van Fraassen 1980; 2001) ground knowledge of reality on human experience and may, therefore, be sceptical of the ontological existence of postulated unobservables (e.g., the "strings" of string theory). The two realist theories mentioned are not intended here to open further discussion regarding the subject of realism (which would fall outside of the scope of the current thesis) but serve as examples of philosophical perspectives as they may relate to the manifest and scientific images.

How should a problem be approached when there are discrepancies between how the world (including ourselves) appears and how it is best scientifically formulated? A rainbow appears to us as a rainbow (a thing), yet it is simply a visual illusion of light refraction (a process). The sun appears to set, yet we know that it is a result of the relative movement of the sun compared to the movement of the earth. A diamond appears to us as solid, yet there is more space on the atomic level than there is matter (Dawkins (2011)[10]. In our ontology, should we accept rainbows, sunsets, and solid objects? These examples reflect reducible explanations, where the manifest image is better explained when reduced to the scientific image. However, there is ongoing contention in the philosophy of mind as to whether the manifest image of the mental (the mind as it appears to us) can be – or should be – reduced to the scientific image of the physical (e.g., Descartes, 1637/1968; Nagel, 1974; McGinn, 1982, 1989). Throughout this thesis, the reliability and the validity of the manifest image is questioned (e.g., sections 2.5.2.1 - 2.5.2.3), whereas the reliability and validity of the scientific image is affirmed. Therefore, the current thesis, while willing to account for the manifest image, postulates from the scientific image as a preferred epistemology.

1.6.8 Causation

For the MUP, the agenda of instantiating a mind in alternative substrates through the use of technology aligns the project with the concept of causation. A discussion of causation is essential to understanding explanations[11] (why certain things happen) and empowers us to predict and adjust to events. According to Schaffer (2016), the standard view of causation is that there is one event and two relata (cause and effect). Causes are related to time, in that the cause (event 1) occurs prior to the effect (event 2).

---

10 Dawkins (2011, pp. 84-85), following the Rutherford/Bohrs atomic model, estimates that the space in a diamond can be conceptualised along the following lines; suppose the nuclei and electrons of the carbon atoms (of which diamonds are constituted) were the size of soccer balls and gnats respectively, then the distance between each soccer ball-sized nucleus (the centre of the atoms) would be approximately 15 kilometres away from each other and the gnats (electrons) would be several kilometres away from these nuclei, and so he concludes that "even the legendary hard diamond is almost entirely empty space" (ibid, p. 85).

11 .Explanations can be seen as identifying and connecting salient causal events. For example, I put the kettle on, which caused the water to boil. In one explanation, a description of how the element heats through electrical current and the properties of water needed to boil and so on, could be included. Another expanded explanation may also include a social cause, such as my wife asking for a cup of tea. However, it would be an absurd and impractical explanation to describe how my parents met, although these events are part of the causal chain (if my father had not met my mother I would not have been born, I would not have married my wife, and therefore this water would not have boiled).

Talbot (n.d.) identifies three primary theories of causation within the philosophical tradition; namely the regularity theory of causation, the counter-factual theory, and the singularist theory. The regularity theory is related to Hume (1777/1975) and asserts that what we deem to be causation are observed regularities (correlations). The counter factual theory, championed by Lewis (1986), relates to the notion of subjunctive conditionals, where we think about what might have been and imagine a scenario where some variable is removed from the process. If the removal of the variable leads to the assertion that the phenomenon would not occur, the variable may be said to be (at least in part) causal. The singularist theory of causation (see Tooley, 1990; Armstrong, 1999) is more closely associated with the empirical scientific account of causation. The assertion here is that the singular (particular) causal is a core notion of understanding causation and that regularities and laws are abstract expressions of singular events.

The precise metaphysical nature of causation falls outside of the scope of the current thesis. All metaphysical theories (regularity, counter-factual, and the singularist theories) would need to account for the observable and repeatable causation evident in all the sciences (including technology development). From a technological point of view, it is, therefore, not essential that this philosophical issue is resolved, but rather that causation is understood sufficiently to conduct and produce the desired technological artifacts that may in turn produce the desired phenomenon.

Within the philosophy of mind, causation (in the general sense) plays a central role in identifying alternative theories about the mind. For example, the problem of mental causation (Robb & Heil, 2019) is the problem of how a non-physical event (whether a non-physical substance or a non-physical property) can cause a physical effect (see section 2.4.1.). Alternatively, if the mental serves no causal role, it is said to be epiphenomenal, such as the steam whistle that accompanies a steam locomotion (it is a non-causal by-product of the system) (Huxley, 1874). For the MUP, it is not essential that the mind itself is causal, but only that the mind (however defined) falls within the purview of some future technology and may, therefore, be caused (artifact production) by scientific means.

1.6.9 Current artificial intelligence (AI) technologies: Relevant considerations

Because much of the mind uploading literature emphasises the computational theory of mind, a further aspect of mind uploading relates to our current thinking on artificial intelligence.[12] Some

12 Not all of the current debates within artificial intelligence is relevant to a metaphysical discussion as they refer to empirical and technical debates. For example, debates over whether AI should follow logic- or probability-based systems (Moor, 2006) do not relate directly to metaphysical mind uploading concerns.

issues that have metaphysical relevance may include, but are not limited to, questions of syntax and semantics, weak and strong AI, as well as plausibility issues such complexity.

The issue of syntax (processes such as grammar) as opposed to semantics (meaning), i.e., how processes such as algorithmic calculations can lead to processes concerning meaning, has been raised by Searle's (1980) work to be discussed in section 2.6.4. Some assert that knowledge of syntax cannot lead to understanding in terms of semantics (e.g., Dreyfus; 1972; Searle, 2004), while others believe that semantic understanding is achievable (e.g., Sloman, 1978; 2001; Thagard, 2017).

The second issue of strong versus weak AI (Searle, 2004) can be seen as the distinction between simulation (weak AI) and emulation (strong AI), as presented in section 1.4. The third issue of complexity enquires as to the limitations of computation and implications for future technology.

In terms of computability, the Church–Turing hypothesis, often a title attributed to the limitations of computation, is attributed to Alonzo Church and Alan Turing. The original idea behind the hypothesis was that all "computable" functions can be performed on a simple Turing machine. However, Church and Turing asserted that not all mathematical problems can be solved in this simple computer (e.g., calculus of the first order predicate is not recursive). Therefore, not all solutions can be sought through computation. Copeland & Shagira (2018) give an overview of the Church–Turing thesis and its many current forms. They conclude that the question of computation is largely open and, with the advent of quantum computing (and potentially other forms), it is difficult to conclude in either direction at this stage.

If the mind were to be deemed a computational entity (whether akin to our current computational systems or something different), then AI and computational theory become increasingly relevant to the discussion of mind uploading. However, if the mind is an entity of both computation and mechanism[13], then mind uploading would need to emulate both the mechanisms and the computational processes.

---

13 Within the context of computation, there are different activities that occur to produce a phenomenon. Consider, a standard computer that can be seen as information processing occurring at the software level, yet at the hardware level there are mechanical activities. For example, the base binary coding (allocation of a 0 or 1) of modern software is allocated by the mechanism of electric conductivity as either 'on' or 'off' (at times called a flip-flop). Early electronic circuits used vacuum tubes as their mechanisms, whereas modern computers use silicon semiconductors (Selkirk, 1995). For mind uploading, this may relate to the mechanism of synaptic firing of neurons (the mechanics of electrical stimulation and conductivity in the brain) and information processing of the mind (e.g., representations).

These considerations will be the backdrop for the two critical foci mentioned above, which will be considered throughout the thesis[14].

1.7 Conclusion

The question of how we can achieve a fully functioning replication of any physical system is an empirical one (an engineering problem). The question of whether a fully functioning replication of a person is in fact a person, and indeed the same person and under what conditions this is maintained, is a philosophical one (the metaphysical problem). The limited current philosophical work related to mind uploading indicates that any enquiry at this stage would be novel. By addressing the two primary premises of, or necessary and sufficient conditions for, mind uploading as asserted by the current thesis (a multiple realizable physicalist metaphysics of mind and a psychological – transferable – personal identity), the current thesis aims to delineate philosophical problems that underpin the feasibility or non-feasibility of the MUP. Through this process of enquiry, it is hoped that new insights can be gained and the thesis aims to map out areas for future enquiry in philosophy as well as potentially for other fields of study of the mind (e.g., psychology, neuroscience, and AI).

Could a particular mind be replicated? In chapter 2, the thesis addresses the concepts of physicalism and multiple realizability as these relate to the MUP. The chapter will explore multiple perspectives in philosophy of mind (including dualism) and query how each of these perspectives would relate to the feasibility, or non-feasibility, of the MUP. The chapter will be orientated within the context of the mind–body problem to better understand what the nature of the mind is and how this nature would relate to the success of the MUP. Each perspective will be evaluated, areas of contention and potential areas for further investigation identifies, and novel contributions will be presented. In essence, the chapter asks what the likely metaphysical nature of the mind is and what this would imply for the possibility of mind uploading.

If a particular mind were to be replicated, would this result in the same person? In chapter 3, the thesis addresses the subject of transferable personal identity (the persistence problem as it relates to the MUP). Again, multiple theories will be considered and evaluated. This chapter enquires into the metaphysical nature of personal identity and relates to the persistence problem and what this implies for the success of the MUP. In chapter 4, the thesis integrates aspects of both the mind–body problem and the persistence problem, as they relate to the metaphysical feasibility of

14 A related issue, which cannot be addressed in this thesis, is whether or not it is ethical to create an artificial entity with a human mind (see for example Harnish & Cummins, 2000). Ethics, however, is not the subject of the current thesis.

the MUP, and develop its own metaphysical stance culminating in a notion of the self. Chapter 5 is the concluding chapter and will summarise the thesis and offer suggestions for further investigation.

2 Towards a multiple realizable physicalism for the mind uploading project

2.1 Introduction

The previous chapter orientated the reader to the mind uploading project (MUP) and the current thesis's objectives of providing a metaphysical stance that would allow for such a project. Various relevant concepts were defined and an overview of some of the current themes within the philosophy of mind in relation to the MUP was offered. It was asserted that the MUP rests on two metaphysical claims, namely, multiple realizable physicalism (in relation to the mind–body problem) and psychological continuity (in relation to the persistence problem of personal identity).

The current chapter focuses on multiple realizable physicalism as a necessary and sufficient metaphysical condition for the MUP. In brief, this asserts that, in order for the MUP to be feasible, the mind needs to be based in a physical (however defined) substrate and that this substrate's properties, and what the substrate does (processes), may be instantiated in an alternative substrate/s (multiple realizability). Both concepts of physical substrate and multiple realizability are to be defined and discussed in more detail throughout the chapter. The instantiation in an alternative numeric substrate is to be an artifact (and, therefore, related to applied sciences) and may, therefore, be defined as a replication/emulation (irrespective of the MUP matrix scenarios, Figure 1-1, section 1.4.).

Artifact engineering requires that the necessary and sufficient properties and processes for replication are understood. The method for artifact engineering is scientific and explores the nature and interactions of physical entities. It is acknowledged that our current understanding of the sciences are limited (including mind sciences such as psychology and neuroscience), with the current thesis using the term 'science' to include current and all possible future scientific developments and relating the term to the setting of relevant nomological boundaries (chapter 1, section 1.5.6).

The current chapter explores what the metaphysical nature of the mind should be in relation to a successful MUP. This investigation is done through the lens of the mind–body problem, which is the primary question in the philosophy of mind that aims to address how physical entities (such as bodies) relate to what appear to be non-physical entities (such as minds). It is argued in this thesis that many divergent philosophical views are consistent with the MUP in so far as each view adheres to a physicalist (in the broadest sense of the term) multiple realizable (in the broadest sense of the term) stance.

There may be multiple avenues to explore the mind–body problem, from historical to thematic approaches. The current thesis opts to first present in section 2.2. the multiple realizability argument. This is because if a mind could not be instantiated across alternative substrates this would nullify the MUP and it is, therefore, imperative that the nature of multiple realizability (as it relates to the MUP) be established at the outset of the chapter. The form of multiple realizability in relation to physicalism and the MUP presents a novel contribution of the current thesis and serves as a lens through which other philosophical stances may be evaluated in relation to the MUP. This section defines the term in relation to the MUP as well as lays the general parameter of what conditions a metaphysical stance would need to take to allow for the feasibility of the MUP.

Prior to engaging with specific philosophical stances to the mind–body problem, section 2.3 presents a brief overview of some of the categories and definitions relevant to the mind–body problem and elucidates the nature of the problem at hand. The aim here is to establish the nature of the mind–body problem and how this problem impacts on the MUP. Furthermore, the current thesis orientates the philosophical literature around the question of what the relevant substrate is as well as around what the substrate does. As stated in chapter 1 (section 1.5.2), this relates to the 'thing' (the substrate) and 'how things hang together' (what the substrate does), with the current thesis emphasising properties of the substrate (what the substrate is) and the processes of the substrate (what the substrate does). This emphasis is a further novel contribution of the current thesis in relating the philosophical literature with the MUP.

Section 2.4 turns to the possibility of the substrate of the mind being non-physical with the emphasis on substance dualism, where the mind is defined as being constituted by a substance separate to the physical substance (the body). In the substance dualist stance, there are two substrates (the mental and the physical), each with their own processes (what the substrate does). The primary criticisms presented against this view are that of mental causation (how can the non-physical substance cause physical activities), over-determination (if physical causes result in observable actions what need is there for mental causes to cause the same observable actions) and the category mistake (confusing entities for their interactions – substrates for substrate processes). Relating to the MUP, if the mind is constituted by a non-physical substrate and artifact design is concerned with replicating physical substrates then the MUP is not feasible.

Following this discussion, a brief introduction to physicalism as the view that acts (in contemporary times) as the primary alternative to substance dualism is offered in section 2.5. Whereas mental causation was shown to be problematic for substance dualism, this is not the case

for physicalism. The emphasis on physical causation places the mind and, by implication, a potential mind replication, within the scope of science. Physicalism, due to its relevance to the MUP, serves then as the primary focus of the current chapter.

Physicalism is broadly defined within the current thesis as having many variations that may relate to the MUP and amounting to a negative hypothesis against the supernatural. The notion of a nomological boundary is used within the chapter to designate necessary and sufficient domains that may be relevant to the MUP should the philosophical perspective under discussion be upheld. The current thesis explores physicalism under two broad categories. The first consists of the views that hold to the physical in terms of standardly recognised physical properties and processes within the current sciences (referred to in the current thesis as standard physical properties and processes). The second category comprises those views that assert that the mind has some properties (which are nevertheless physical) that are beyond what is currently scientifically established physical properties, i.e., 'further physical' properties as defined in chapter 1. In relation to substrates and what the substrate does, the standard physical substrate is taken to be constituted by mass, energy, biochemical reactions, and so on. The mind is largely viewed as a complex process (what the substrate does with these standard physical properties). The views concerning further physical substrate make the assertion that the mind not only includes these standard physical properties but has some further physical property/ies and, therefore, asserts property dualism (see sections 2.7).

The chapter first focuses on physicalist views based on standard physical substrates and categorises these further as to those who emphasise what the substrate does in the environment or externally (e.g., behaviourism – sections 2.5.1.1.) and what the substrate does internally (e.g., identity theorists – sections 2.5.2.1.). These categories are termed views based on 'external processes' and views based on 'internal processes', respectively. Those who question the standard physical properties views are discussed in relation to Nagel's (1974) *What it is like to be a bat* and Jackson's (1982, 1986) knowledge argument. Because those who question standard physical substrates nevertheless retain physical monism, while arguing for an expansion of further properties (mental properties), these views may be aligned with property dualism (see section 2.7).

The thesis then introduces the approach known as functionalism, which allows possibilities of views based on standard physical substrates as well as those based on further physical substrates and, therefore, acts as bridge in the thesis between these two physical monist stances (the monist who holds to standard physical substrates and the monist who asserts further physical substrates). Different forms of functionalism are presented together with explanations of the notion of emergence (both weak and strong) as being congruent with various functionalist perspectives. For

example, if strong emergence within functionalism is asserted, then it could be said that a further physical substrate (a new property has emerged – emergent properties) is present and a necessary (although perhaps not sufficient) condition for the mind. Common critiques of functionalism are also addressed here.

Section 2.7 next considers further physical substrate views under property dualism (the assertion that the mind is constituted by mental properties that are beyond standard physical properties). This section evaluates the works of Chalmers, Kripke, and Davidson in this regard. As in previous sections, these arguments are briefly critiqued, however it is acknowledged that the MUP is also defended if property dualism is upheld. It will be shown that property dualism is consistent with multiple realizability and, therefore, the MUP.

The final section summarises the chapter, identifying salient themes as they relate to the mind–body problem and the MUP.

2.2 Multiple realizability physicalism and the MUP

The current thesis has put forward multiple realizable physicalism as one of the necessary and sufficient conditions for the MUP, with the current chapter emphasising that the substrate and what the substrate does are to be physical (however defined). 'Realizability' is used throughout the thesis as synonymous with 'instantiation' because both concepts assert that a phenomenon occurs (is instantiated, is realized). For a phenomenon to be multiply realizable/instantiated it would, therefore, need to be of a nature that it could occur more than once (multiple). For the MUP, this relates to the idea that a specific mind may be realized/instantiated in an alternative substrate (e.g., a robot or clone) other than the original substrate (e.g., the current biological body).

Realizability/instantiation, as is broadly described above, may occur at multiple levels. Consider the carpenter who is making a table. The phenomenon of solidity occurs in multiple mediums (e.g., different woods) and, therefore, this phenomenon is multiple realizable, at this nomological level. The atoms (things) interact (processes) occur at the atomic nomological level and the phenomenon of solidity is realized.[15] However, the multiple realizable phenomenon of solidity is not sufficient for the phenomenon of a table (e.g., a diamond ring is solid but is not a table) and the

15 Therefore, the nomological boundary accepts that each nomological domain puts forward things and how they hang together. For example, atoms (things) interact (processes) to form molecules (the realized phenomenon), molecules (things) interact (processes) to form organisms (the realized phenomenon), and so on. At each nomological level there is therefore realization of a phenomenon and the question for the MUP is whether the phenomenon of a specific mind is multiple realizable.

woods need to be joined in such a way (a higher level of interaction of the solidity phenomenon) to produce the table. The phenomenon of a table is also multiple realizable (e.g., the carpenter may produce many tables) and is an artifact replication (technological replication). Multiple realizability may, therefore, occur at multiple levels and relates to things (e.g., wood) and how things hang together (the table design).

The above examples also relate replication to multiple realizability, in that any phenomenon that may be replicated (e.g., tables) may be said to be multiple realizable (e.g., many tables may be replicated and, therefore, the table phenomenon is a multiply realizable phenomenon). In essence, in this sense, what is replicable is multiply realizable. If one considers replication as an indicator of multiple realizability (whatever may be replicated occurs in two space-time locations), one property appears to be non-multiple realizable, at the level of everyday objects, that of the space-time continuity of the substrate. In the table example, the same table design phenomenon occurs (is instantiated, is realized) but the two substrates (the two tables) occupy different space-time continuity. Whether this one property of substrate space-time continuity is what matters to minds and persons is the subject of chapter 3 (particularly section 3.5.).

In this generous interpretation of the concept of multiple realizability, the abstract concept of mind is clearly a multiple realizable phenomenon. The author and the reader both have minds and so the phenomenon of mind is realized in multiple instances. The question for the MUP is whether a specific mind (e.g., your mind, my mind) is a multiple realizable phenomenon. Furthermore, for the MUP, this specific mind needs to be replicated (artifact creation). From a metaphysical (as opposed to engineering) perspective, the mind would, therefore, need to be of a multiple realizable nature (it could, in principle, occur across substrates in some possible worlds), whereas whether this is practically feasible (the ability to build such an artifact) is a problem for engineering. It will be demonstrated throughout the chapter that multiple realizability (as it applies to the MUP and as defined here) is not limited to one form of physicalism but many forms and it is acknowledged here that this may broaden the notion of multiple realizability beyond what has been traditionally used (see section 2.6.3) within the philosophical literature, to which the current thesis now turns.

2.2.1 Defining multiple realizability

The primary emphasis of the term multiple realizability, as it is used within philosophy of mind (Audi, 1999; Bickle, 2019), relates to the same mental phenomenon occurring across different substrate kinds (i.e., the substrates have variation of physical properties). For example, consider

whether the same mental phenomenon such as pain could be realized across multiple substrate kinds such as animals (e.g., octopus), extra-terrestrials (e.g., Martians), or non-biologicals (e.g., computers). Each of these substrates is a different kind than the human kind, yet it can be argued that they are all likely to have the same mental phenomenon of pain. For Endicott (1993; see also Kim 1992), this has led to discussion on whether each mental phenomenon may be causally species-specific. This species-specific view would attribute pain to different species (e.g., an octopus) but deny that it is the same mental phenomenon (it has different physical properties and causal processes) as human pain. However, the species view also asserts that the same mental phenomenon (e.g., pain) occurs across different numeric substrates of the same kind (i.e., the mental phenomenon of pain is realized in different numeric substrates of the same substrate kind).

Whether one holds to the same substrate kinds as necessary and sufficient for multiple realizability or not, does not impact on the assertion that mental phenomena are not restricted to a particular substrate. If it is assumed that multiple realizability may be realized across different substrate kinds, then mental phenomena may possibly (depending on empirical restrictions) occur within non-biological substrates such as computers; and option A (artificial) artifacts of the ABC options (see section 1.3.) may be a potential avenue to pursue for the MUP. If the alternative is assumed, such as species specific properties, then phenomena only occur within the same substrate kinds and the MUP would need to pursue a biotechnological solution (option B of the ABC options), such as a DNA clone, that could realize the same mental phenomena.

Another area of distinction (e.g., Block, 1980; Horgan, 1984; Bickle, 1992, 1996; 2019 section 1.5) of multiple realizability relates to whether the same mental phenomenon may occur (is instantiated, is realized) within the same numeric substrate. For ease of discussion within the current section, consider a particular individual person (e.g., yourself) and the distinct biological activities that occur within the overall substrate (e.g., different neurons and their activities that occur within your body). For example, consider the mental phenomenon of a sharp pain experienced from stubbing a toe on a table in the middle of the night. If one were to stub the same toe on a different night, it could be said to be the same mental phenomenon of the pain of stubbing one's toe, despite some variation in physical activity (e.g., the precise location on the toe, the precise neural activity). Because the same mental phenomenon (the pain of stubbing a toe) is attributed to multiple physical realizations (variation of physical substrate components and their activities) within the same overall substrate (e.g., the body), it can be said that the mental phenomenon is multiple realizable. If there can be multiple realization of mental phenomena in the same substrate, then there is the possibility that the phenomenon could occur in an alternative substrate, if the appropriate conditions are met.

What is important for the MUP is whether the mental phenomenon may occur in an alternative numeric substrate and, therefore, the mental phenomenon cannot be bound to a particular substrate, hence it will be multiple realizable. Gillet (2002) presents two forms of realization namely, flat realization, where the realizer properties and the realized can only occur in the same individual (i.e., the same numeric substrate); or dimension realization, where the realized and the realizer properties may occur in both the same individual or within other individuals (i.e., may occur within different numeric substrates). The current thesis, therefore, aligns the form of multiple realizability required by the MUP to dimensional realizability.

Flat realization faces various difficulties. For example, if a particular phenomenon can be instantiated in a particular substrate (as is asserted in this view), then it may be possible that it can be instantiated in another (e.g., the story phenomenon in multiple books, the statue phenomenon in multiple mediums, pain instantiated in different persons, and so on). To negate this, the phenomenon would have to be uniquely bound to a particular substrate (the further assertion of this view). How would this binding occur? The current thesis puts forward that the flat realizer is left with either asserting that the substrate has unique properties or that it has unique processes.

The current thesis presents the difficulties of a unique phenomenon as the uniqueness problem, where uniqueness takes an absolute form (it cannot be multiply realized under any circumstances). If the substrate has unique properties, then this particular substrate would be constituted by properties that do not exist elsewhere in the universe (if not, the properties would not be unique). Given our understanding of the similarities of standard physical properties within the known universe (here assuming physical monism), it appears that unique properties for each mind's substrate would be unlikely. Note that this would require that the properties be significantly unique so that the phenomenon could not be repeated in any alternative substrate (i.e., minor variations, such as different compositions of clay making the same type statue, are not significant to negate the multiple realization across numeric substrates). If a unique process is posited, this refers to an issue of complexity, as the same (or similar enough) properties are available but the particular interaction is too unique for replication. The difficulty of complexity may contra-indicate replication (and, therefore, multiple realizability), but it is noted that this is a difficulty of pragmatic engineering and not metaphysical difficulty. If it is assumed that the mind is non-replicable because it is too complex, it is similarly not a metaphysical claim but an engineering one. If the assertion of complexity is given metaphysical status it would need to assert that, despite similar properties, the interaction of these properties could not be repeated in any possible world under any possible circumstances and validation would have to be provided for such a claim. Therefore, flat realization

appears less likely than dimensional realization in relation to the mind; however, if flat realization were to be upheld, then the MUP would not be feasible.

The current thesis, therefore, defines the term multiple realizability with different emphasis along a spectrum from instantiations occurring only in a particular numeric substrate (i.e., each individual person) to occurring across multiple kinds of substrates. If multiple realizability refers to instantiations that only occur in one physical substrate (a numeric person), then, if this is upheld, the MUP is not feasible as only one substrate can be said to instantiate a particular cluster of mental phenomena (flat realization). If multiple realizability refers to that which occurs across substrate kinds (e.g., humans and computers), then the MUP may pursue either the A or the B of the ABC options as this assertion. If the form of multiple realizable across substrate kinds is denied (e.g., species specific realizers) but the view of multiple realizability is still affirmed across alternative numeric substrates (e.g., the mental to occur across substrates of the same biological constitution), then the B option for the MUP is preferred. As mind uploading only requires that the mind be realized in an alternative substrate (irrespective of kind as indicated in both A and B options), a broader interpretation of multiple realizability is acceptable. Therefore, the current thesis adopts a broad notion of multiple realizability to include any realization that occurs across a distinct numeric substrate (whether it be of the same or a different kind).

A related term that may be distinguished from multiple realizability but also indicates multiple realizability or is compatible with it, is that of supervenience (see Davidson, 2001; Kripke, 1972, Kim, 2010; Horgan, 1993). In its simplest form, supervenience may be stated as the assertion that "there cannot be an A-difference without a B-difference" (McLaughlin & Bennett, 2018, para. 1), with variation of strictness in relation to the term 'cannot' (e.g., 'cannot' may be seen as a metaphysical or a nomological claim). Within the philosophy of mind, 'A' may be associated with the standard physical processes and 'B' with the mental. Supervenience is also associated with token physicalism (Davidson, 2001; see sections 2.5.2.1 and 2.7.3), in that various physical events (variation in neural activity) may be seen as multiple tokens of the same mental type (e.g., a particular pain). In this case the mental (type) supervenes on the physical (token/s), in that there can be no mental change without physical change (e.g., the neural correlate to the mental event). Whether at a supervenient level the mental results constitute new properties (section 2.7) or whether this level indicates a variation of description of complex, standard physical processes (e.g. Dennett, 1989, see section 2.5.2.4) is one of the themes (developed between standard physical and further physical substrates) within the current thesis.

McLaughlin and Bennett (2018) use an example to illustrate the basic principle of supervenience in considering a perfect forgery of a painting. The micro-physical properties (colours, shapes, and so on) are precisely the same and the image is said to supervene on these. Any change in the image (i.e., moving an object within the image) would result in changes on the micro-physical properties (and *vice versa*). Therefore, if there are the same micro-properties (occurring at a lower level, such as colours), the supervenience principle asserts that the same higher level phenomenon will be present. In the above painting example, not all conditions need be the same for the same phenomenon to supervene. For example, the supervenience of the painting image is upheld if the same lower levels (e.g., colours) are maintained, yet the condition of substrate continuity in space-time is different (the original painting may be in one place and the forgery in another). Therefore, supervenience of a phenomenon (e.g., the image) is not dependent on all conditions remaining the same and, in particular, is not dependent on the continuity of the same substrate (the image supervenes on both the original and the forgery). Supervenience is, therefore, related to a type (a phenomenon) with multiple tokens (multiple realizations).

Furthermore, that not all conditions are needed to be maintained for supervenience to occur may be seen when considering that within the original painting there are variation of conditions at lower nomological levels, such as quantum variation. Within the quantum level of everyday objects, such as paintings, there is constant quantum fluctuations (Serway *et al*., 2018) and, therefore, not all conditions (i.e., all quantum phenomena) are sustained in the original painting. The notion of a nomological boundary may assist by stating that the necessary and sufficient level for the painting is that of colours and how they hang together (in which case, the image may also supervene in multiple mediums, such as a painted canvas, a computer screen, a photograph, and so on) and to ignore the quantum level. The supervenience principle may, therefore, apply to artifact replication by asserting that the necessary and sufficient conditions are a) identifying whatever nomological boundary is appropriate and b) replicating these conditions in an alternative substrate. In relation to the MUP, the question is what is the appropriate nomological boundary of minds/persons, and whether these conditions are replicable such that the same mind/person may supervene in an alternative substrate.

Multiple realizability may also been seen as a feature of nature as presented within the context of the scientific method (Born, 1949; Popper, 1983; Leek and Peng, 2015). The method includes creating or adopting a hypothesis (tested with empirical observations/experiments) that gains credibility (verifiable, falsifiable, and so on) if the experiments are repeatable with the same results (replication or reproducibility). If it is assumed that the construction of a scientific theory

requires the repetition of experiments with numerically different equipment and materials, it can be said that natural phenomena are, in principle, multiple realizable in the context of replication of experiments (i.e., the scientific theory posits a property or process – the observed effect of the experiment – that, if the theory is correct, will occur in alternative substrates and locations). For example, if mass can be measured in two locations, using two different substrates, the phenomenon of mass may be said to be multiple realizable (the phenomenon of mass occurs across substrates and locations). The term replication is used in the context of the experimental method but note that technological replication refers to the application of scientific knowledge such that a desired phenomenon may be instantiated/realized through an artifact/s (example in next section). The point of the current chapter is merely to assert that the nature of the scientific method is to assume that natural phenomena that are experimentally investigated or (re)produced tend to occur in multiple locations and across multiple substrates and that this may bolster the generous view of multiple realizability emphasised in the current thesis.

Taken in its broadest possible terms, multiple realizability is defined here as any realization/instantiation, at whatever nomological level, where the same phenomenon is realized across a distinct numeric substrate (i.e., a phenomenon that may occur in different space-time instantiations in alternative substrates). If the phenomenon cannot be realized across distinct numeric substrates, then it cannot be said to be multiple realizable. This generous view of multiple realizability is governed by the needs of the MUP, where the mind (however defined) needs to be realized in an alternative substrate (the substrate of the upload) from the original substrate (the original biological substrate) to be deemed a successful upload of the mind.

2.2.2 The argument for multiple realizable physicalism

The current thesis asserts that multiple realizable physicalism may serve as common ground for physicalist theories that affirm the MUP. It will be demonstrated throughout the chapter that many physicalist positions here may include multiple realizability of some sorts. The basic argument for multiple realizability within physicalism can be described as follows:

(1) The mind is a physical process or properties (however defined).
(2) Physical processes and properties are likely to be multiple realizable (as defined above).
(3) Therefore, the mind is likely multiple realizable.

The first premise (1) distinguishes physicalism from substance dualism. The emphasis here can be seen as a negative hypothesis, in that it denies any form of supernatural (beyond nature) entities. It could be described as a naturalist stance, in that it affirms only the natural (physical) and follows the physicalist assertion that all that is ontologically real is physical. A distinction between the mental and the physical will be shown to largely rest on intuition based on the notion of the manifest image and/or category mistakes. The primary category mistake that will be emphasised within the current chapter is that of mistaking things (entities) for how things hang together (interactive processes), and most physicalist stances acknowledge that it would be a category mistake to confuse entities for processes. The first premise does not make distinctions about which philosophical physicalist variation is preferred (e.g., identity theorist, functionalist, property dualist), but simply asserts that the mind (however defined) is physical (however defined).

The second premise (2) is a restatement of multiple realizability as presented above and asserts that physical properties and processes are likely multiple realizable. One reason for this assumption is that known physical properties and processes are multiple realizable. Consider any physical property that occurs within the domain of everyday human experience (e.g., solidity, temperature, speed, energy, mass, volume) and it is noted that these properties and processes are multiple realizable (both across substrate kinds and across numeric substrates). Furthermore, the empirical trajectory (see section 1.5) of modern science and technology indicates the probability of replication. Within artifact creation (technological replication), a phenomenon may be said to be repeating the circumstances under which certain necessary and sufficient conditions are re-instantiated in an alternative substrate/s. For example, consider the phenomenon of electrical discharge, which in nature may occur as lightning. In the artifact (e.g., a Tesla coil), the phenomenon of electric discharge may be repeated and, therefore, replicated (in this sense). Technological replication is, therefore, the application of scientific knowledge such that a desired phenomenon may be instantiated/realized through an artifact/s.

If this replication, in this sense, can be performed with standard physical properties and processes (the current domain of science and technology), then the empirical trajectory indicates that any physical property (e.g., one to be discovered in the future) is also likely to be replicable. The assertion that the mind is somehow related to further physical properties and processes (see section 2.7) would not negate this probability and, therefore, also likely imply the mind to be multiple realizable. The implication of the empirical trajectory is that as science (broadly defined) progresses, more and more properties and processes are discovered and that each fall within the pur-

view of scientific knowledge and an increasing number of observable effects are repeatable (technological replication). That the mind will be within the purview of science is not, however, without it critics and the current thesis addresses two primary criticisms from Nagel and Jackson in section 2.5.3.

The multiple realizability argument presented here is not a necessary logical argument, but a probable contingent argument. Simply stated, given that most known physical phenomena are multiple realizable (along with the ongoing success and trajectory of increased knowledge in the sciences and replication of phenomena through technological artifacts), it is likely that there will come a time in the future when the mind, as a physical system (irrespective of whether a standard physical or further physical substrate), will itself be replicable.

This is not to say that all physical phenomena are necessarily replicable. For example, the no-cloning theorem (Park, 1970; Daffertshofer, Plastino & Plastino, 2002) argues that certain quantum states are impossible to clone. Although solutions for sufficient replication have been put forward in relation to the no-cloning theorem (e.g., Bužek & Hillery, 1996), the current thesis acknowledges the possibility that certain physical properties and processes do not allow for replication. However, it is noted that examples of physical properties and processes that may not be replicable do not occur at the biological level (the level at issue in the context of the mind–body problem) and the relevance may, therefore, be queried. The current thesis, therefore, acknowledges that mental phenomena, even if physical (however defined), may possibly be one such physical phenomenon and that would contraindicate the MUP (C of the ABC options). However, this is deemed an improbable outcome for the reasons given above and would await further empirical research.

Figure 2-1 broadly illustrates the options within the mind–body problem (serving as the general structure of the thesis's approach to the problem) as they relate to multiple realizability and the MUP.

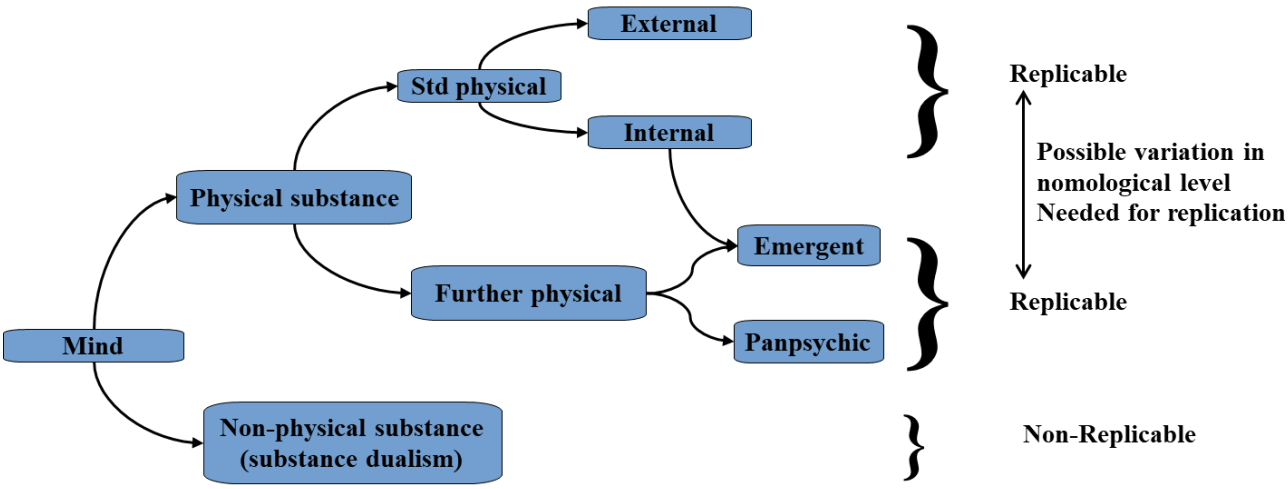

Figure 2-1: Mind replication and the mind–body problem

The first distinction in the above diagram is between the mind as a physical or a non-physical (with the focus in this thesis on substance dualism) phenomenon. This relates to what substance/substrate is related to the mind and is the primary focus of the mind–body problem. A non-physical substance is not replicable within the MUP, as the project falls within the purview of applied sciences and, therefore, relates to physical (however defined) substances. The physicalist stances can then be broadly distinguished between those that assert the mind is instantiated in a substrate that has standard physical properties and processes, or those that assert the mind is instantiated within a substrate with some further physical properties or processes.

Within the current thesis, the standard physical substrate stances are further categorised in terms of a focus on external processes and an emphasis on internal processes. A variation of the emphasis on internal processes that integrates the possibility of further physical properties is the emergent view (whether strong or weak emergence is held see sections 1.5.6 and 2.7.4). In terms of emergence, the standard physical substrate and its processes interact to form a new property (e.g., solidity or liquidity emerging through atomic and molecular interactions as lower levels). In essence, the higher level property emerges from the lower level interactions (things interact to form new things). In this way, the emergent property is dependent (supervenient) on the lower level. As standard physical properties are commonly replicable, those who confirm standard physical substrates and their processes (whether this leads to emergence or not) as the basis of mental interactions, are likely to allow for the MUP, with an emphasis on the lower nomological level replication

(e.g. a biological artifact with the same neurobiological properties and processes) being necessary and sufficient for the emergence of the higher emergent properties.

If the mind is asserted to be some further physical property or processes beyond the standard physical phenomena, then either the mind is an emergent property (emerging from the lower standard physical processes) or it is a further property that is concurrent with standard physical phenomena (panpsychism). In this view, these concurrent further physical properties are what the mind is constituted by and, therefore, to replicate the mind, these panpsychic further properties and processes are necessary and sufficient. To replicate a mind, in this view, would then not require the lower nomological levels (e.g., neurobiology), but rather an emphasis on the higher nomological level (e.g., psychology, here defined by further panpsychic properties and processes) would be necessary and sufficient for the MUP.

Figure 2-1 represents the structure of the current chapter, with section 2.4 relating to the non-physical substance, section 2.5 relating to the standard physical view, and section 2.7 relating to the further physical view (section 2.6 relates to functionalism that may traverse the standard physical and further physical substrate views). Despite the variation of views held within the mind–body problem, it will be shown that multiple realizable physicalism (as defined within this thesis) is consistent with all forms of monist physical substance perspectives evaluated. Furthermore, it will be shown that multiple realizable physicalism allows for the metaphysical probability of the MUP in relation to the replication of a particular mind. Replication, however, raises the question of whether the sufficiently replicated mind can be designated to be the same person and how this replication would relate to the persistence problem (that persons persist yet change over time). This is the subject of chapter 3.

2.3 Mind-body distinctions

The distinction between the mind and body is referred to in the literature as the mental and the physical, respectively. The physical domain is within the realm of the natural sciences (with the emphasis on causation), whereas the mental may be defined as "pre-scientific ways of describing and explaining human behaviour" (Jaworski, 2011, p. 23). This manifest image of the mind, as the current section elucidates, has multiple features and may even be viewed in terms of folk psychology (the common everyday understanding of the mind as a meaningfully causal process). McGinn (1982) presents the problem as that when "we think reflectively" (p. 16), there appear to be "two sets of truths" (p 17). The first 'truth' is that the mind appears embodied in the physical body in some way (i.e. it is physical) and the second 'truth' is that the mind appears to have non-physical

properties, such as first-person perspective, intentionality, consciousness (i.e. it is non-physical). Therefore, there is a tension between two manifestations of the mind, which has both physical and non-physical properties. The mind–body problem simplified may be formulated as follows:

(1) The mind has mental features, e.g., desires, beliefs.
(2) The body has physical features, e.g., occupying particular space-time, energy relocation.
(3) How can these two disparate features be reconciled? (the problem)

The current thesis emphasises causation in relation to the mind–body problem firstly because causation is commonly used to evaluate various philosophical stances (e.g., sections 2.4 and 2.5), but also because causation is essential to artifact creation (e.g., knowledge of how substrates causally interact internally and externally). Broadly speaking, for those who hold to a substance dualism, the problem can be formulated as how a non-physical entity (the substance dualist definition of the mind) can have causal efficacy with the physical (the body) and *vice versa* (how the physical body can cause changes in the non-physical mind). Whereas for those who hold to a physicalist stance, the problem can be formulated as to how mental features can extend from (be caused by) physical features. Furthermore, are mental features themselves causal or epiphenomenal, or are mental features even real or a manifest illusion? For the physicalist, the standard physical properties and processes are largely understood as causal, which leaves different options of how the mental relates to these properties and processes (presented in each section under physicalism and integrated in section 4.1.1 of chapter 4). For example, the mental could be understood as being causal because the mental is merely an alternative description of standard physical processes (section 2.5.2.1.), or the mental can be a causally distinct physical property from standard physical properties (section 2.7.1). However the physical relates to the mental the issue of how to account for causation is an important emphasis, particularly in relation to artifact creation, where the causal properties and processes are replicated in an alternative substrate.

Because physicalism asserts that there is no non-physical mental events or entities (all reality is physical), it may be misleading to distinguish the mental from the physical in the traditional manner (allocating the physical in contrast to the mental) because this may appear to concede that there are two distinct substances (the very assertion that is denied). As it would be a misnomer, from a physicalist perspective, to contrast the mental with the physical (as both are asserted to in

some sense be physical), the current thesis prefers to speak of higher and lower levels[16] of physical interaction, with the higher being associated with the mind (psychology) and the lower with the body (biology). Throughout this thesis, the higher levels are designated as Hl and the lower levels are designated as Ll and relate to the mind (Hl) and the body (Ll), unless otherwise stated. The current chapter, which emphasises physical monism, explores different philosophical stances and how these views perceive the relation between mental (Hl) and physical (Ll) as well as how these views offer potential solutions to the mind–body problem in relation to the MUP. At present, this section identifies areas of some distinctions that are presented within the philosophical literature between the mind and the body. A common way to distinguish between the mind (mental) and the body (physical), is to view the mental from a first person, subjective, and internal perspective, whereas the physical is observable from a third person, objective, and external perspective. First person perspectives are, therefore, introspective, as the mind's eye turns in on itself and makes observations on the organism itself. This first person observation is believed to offer perspectives and experiences that cannot be achieved through the third person perspective and, therefore, is a distinct epistemological stance and is, in some sense, privileged information. Historically, a person can retain thoughts that are not subject to third person observation (third person inaccessibility) and, therefore, are deemed to be private and privileged (related to terms such as authority, incorrigibility, and infallibility). However, the progress within neuroscience implies that much of what we thought to be private may yet be accessed by third persons. For example, Wen *et al*. (2018) have been able to decode neural activity and reconstruct moving images (third person objective) perceived by the subject (first person subjective). In this sense, privileged perspective of the first person subjective may be challenged and a scientific objective third person perspective would then likely offer a more reliable and valid account of the subjective experience (based on the rigours of the scientific method in comparison to the *post hoc* method of personal psychological development).

A further way to distinguish the mental from the physical is to focus attention on the subjective feeling, which may be termed qualia, phenomenal consciousness, what-its-like, and so on

16 Although notions of levels may imply emergence (whether weak or strong), this need not always be so. Higher levels in this thesis, unless otherwise specified, are a description of the psychological and, therefore, may be identified with lower levels (identity theory – section 2.5.2.1), or run parallel to other physical substrates (panpsychism – section 2.7.1.) or any other such alternative, such as supervening on the lower levels, as discussed throughout the thesis. The choice to prefer the term 'higher levels' to 'mental' in this context is, as stated above, to avoid an assumption of distinction between the physical and the mental at the outset of the enquiry.

(e.g., see Nagel, 1974; Block 1995, 1996a, 2008). Qualia are notoriously difficult to define but are generally understood as the properties of mental states or events that relate to sensations or perceptions that lead to "what it is like" (from here on abbreviated to WIL), feelings, raw feelings, phenomenal properties, or qualitative experiences such as itching, colour perceptions, taste, pain, and so on. This subject is further evaluated in section 2.5.3, which addresses, primarily, the argument from Nagel (1974). For Jaworski (2011), a standard reading of qualia (and phenomenal consciousness) is that it is intrinsic (non-relational) and simple (unanalysable), meaning that within the first person experience (the manifest image) they cannot be de-constructed but appear to us as a whole. When we experience itches, colours, pain, and so on, they appear (manifest) to us as indivisible entities and, therefore, it may be inferred that what atoms are to the physical domain of everyday life (here de-emphasising the subatomic theories which are also part of the physical domain), qualia may be to the first person experience. However, just as the atom (derived from the Greek meaning to be unable to cut or divide) was divided in recent times, and theoretical knowledge has reduced it to more elemental particles, so too the quale may be reduced to neurological features or defined as illusion, or eliminated from meaningful speech (Dennett, 1988a).

The mind may also be seen as having intentional features (Brentano, 1995; Huemer 2019), which, at times, are termed propositional attitudes. Intentionality is a form of representation capacity and function, in that it is a state that is about something. This "aboutness", according to this view, is the fundamental essence of mentality that is directed towards objects (see also McLaughlin, 1991 and Searle 2004, Jacob 2019). One type of intentionality is propositional attitudes (content intentionality) that are psychological states usually expressed in language, where the verb is related to the individual enacting an attitude toward particular content. I *believe* it is raining. I *hope* to achieve good grades. I *want* to eat the cake. In each case, there is a proposition "it is raining; achieve good grades; eat the cake" and these are related to an attitude of the individual to the proposition and expresses their intentionality. Furthermore, the mind is rational (Davidson, 2001). Even in psychosis, an underlying rational is assumed. For example, a person suffering from delusions of grandeur may believe they are immensely important (e.g., the king of England), despite their current state of poverty and low status. A psycho-dynamic (that which explores the underlying unconscious mental dynamics) may assume that this delusional belief is a response to protect a diminished sense of self-worth. Therefore, it assumes an unconscious rationality to the conscious irrational delusion.

Given that the mind has these features and the physical appears not to (e.g. mass has no rationality, energy has no intentionality, weight has no qualia, and so on), solutions to the mind–body problem lie in how to best understand this disparity. This difficulty in explaining how these

two domains of enquiry (the psychological and the physical) interact has been described in terms of the 'explanatory gap' (Levine, 1983). It is asserted that there is a gap between explanations provided by scientific knowledge of the physical (e.g., here meaning causal functions) and the experience (feeling) and processes of the mind (ibid.). For the substance dualist, the disparity is explained by asserting that the mind is of a different substance. For the physicalists, two primary options have been presented within the current thesis as views based on the standard physical substrate and views based on the further physical substrate. If the gap between the physical and the mental features is insurmountable, yet physical monism is upheld, the mental features must refer to some further physical property and, therefore, the distinction must be taken as foundational to the debate (e.g., Chalmers 1995; 1996). For others (e.g., identity theorists, eliminativists, reductive physicalists), the gap is either reducible to physical states and events or is superfluous (e.g., Dennett, 1988a; 1991). In the latter sense, the gap appears as either a matter of misunderstanding (e.g., the appearance of qualia are an illusion) or is a matter of complexity (the qualia may be reduced to standard physical processes).

Let us consider the implications if the gap is affirmed (the mind as non-reducible) or denied (the mind is reducible). If the mind is non-reducible (the gap is affirmed), yet it is physical (as affirmed by physical monism), what is required for the MUP is to replicate that (higher) level (e.g., a representational mind requires replication of the representational system alone). Because the mind is deemed to be constituted by things (e.g., representations) and how they interact (e.g. representational processes) and these are non-reducible, then replication of these things and how they hang together at this nomological boundary will be necessary and sufficient for the MUP. If the gap is denied, then replication of the lower nomological domains (e.g., neuroscience), the higher nomological domain of psychology (similar to the carpenter who works in the autonomous higher domain of woodwork despite it being reducible to atomic theory), or a combination of both higher and lower domains may be necessary and sufficient for replication, The relation between the Hl domains and Ll domains are the subject of further discussion throughout the chapter, and what is argued here is that whether the gap is affirmed or denied need not negate replication but limit the necessary and sufficient nomological boundaries.

This section has presented some of the relatively uncontested features of the mental to assist the research. One perspective that orientates these features is a general mental model, such as the Desire-Belief-Action Principle (DBA). This model is defined by Kim (2011, p. 73) as "If a person desires that *p* and believes that by doing A is an optimal way to secure *p,* she will do A". This can also be adapted to state that she will be disposed to do A absent defeaters. It should be

noted that within any scenario there may be competing desires, but for the sake of argument, a singular desire without conflicting desires is assumed here, such as Liezel desires to drink some water. She leaves her chair to pour a glass of water and drinks the water. In this sense, DBA provides an explanation (a pre-scientific manifest image) of what casual activities occur within the mental domain that results in action. It is assumed that these desires and beliefs are first-person, intentional, consciousness, rational, and are likely to include qualia, WIL, and so on.

This section has not presented an exhaustive overview but emphasised certain features that are commonly used to distinguish the mind from the body. The mind may be described as having mental features, such as first person subjective perspectives (including authority and privacy), first person experiences (e.g., WIL, qualia), desires, beliefs, intentionality, propositional attitudes, and rationality. The body may be described as having standard physical features, such as being third person observable, and so on. The thesis will continue to explore these aspects as they are discussed within various philosophies of mind.

2.4 Non-physical substances

This section focuses on philosophies of mind that assert the mind to be, in some sense, non-physical. Within this thesis, 'non-physical' refers to the substance/substrate, where properties of the mind are instantiated and which can exist independently of the physical. 'Substance', therefore, relates to the substrate, where the mind is instantiated. Within these theories, asserting that the mind is non-physical, substance dualism offers the greatest resistance to the MUP, in that the project requires a human artifact to replicate a mind, and if the mind is constituted from non-physical substances this would be beyond the purview of the project. Prior to addressing substance dualism some alternative non-physical philosophy of mind perspectives are presented.

The non-physical mind can be seen on a spectrum from monist idealism to dualism (in its various forms). Monist idealism (e.g., Berkeley, 1713/2017; 1734/2002) asserts that all of reality is non-physical and matter (the physical) is denied. Occasionalism (Lee, 2016) asserts that the mind and the body are separate entities but that the only true cause in the universe is God. Parallelism (ibid.) asserts that the mind and the body are separate entities but that there is no interaction between these entities and any correlation that may be observed between activities of the mind and activities of the body are causally disconnected. A further form of parallelism, according to Bobro (2017), is pre-established harmony (see Leibniz, 1646–1716), where the mental and the physical both have no causal effects but are pre-established by God to co-occur in harmony, such as two clocks set to co-occur in time. The current thesis now turns to the substance dualism.

2.4.1 Substance dualism

This section deals with interactionalist (or substance) dualism, where the mind is thought to be a separate substance to the physical yet the two interact to some extent. Interaction between the mental and physical can, therefore, be formulated as follows M ←→ P. The mental can thus cause physical events such as in M→P, where a purely mental event (without a physical antecedent) can cause physical outcomes as well as some physical causes having no effect on the mind. It is this aspect of dualism, M→P, that relates to the MUP as it assumes a mental cause that cannot be replicated through physical artifacts.

Descartes' texts (1637/1968) serve as the seminal text on what is now termed Cartesian dualism, which has been the subject of continuous debate within philosophy. Summarised aspects of this view, as they are understood in the literature and that reflect aspects relevant to the MUP[17], are presented below. For Descartes (and modern substance dualists), the mind–body problem is maintained primarily through some form of Leibniz's law, in asserting that the mind has certain features that the body does not and from this disparity of features it is then deduced that the mind is of a different substance to that of the body. For Descartes (ibid.), the primary emphasis is that the mind is a thinking entity (*res cogitans*) and a non-extended entity. In relation to the notion of a thinking entity, Descartes states, "I rightly conclude that my essence consists only in my being a thinking thing [or a substance whose whole essence or nature is merely thinking]" (ibid, p. 132). That one of the functions of a mind is thought (however defined) is not contested and can be accepted as a feature of the mind. However, that thinking may be used as a distinction between a thinking substance (the mind) and a non-thinking substance (the physical), as Descartes asserts, has been contested by physicalist philosophers (see section 2.5).

Furthermore, the mind, according to Descartes (1637/1968), is non-extended where the notion of extension refers to the ability of an entity to occupy space that is the domain of the physical, whereas, in contrast, the mind occupies no space (cannot be extended) and is, therefore, a non-physical entity. For Descartes, the physical, as extended, allows for the physical domain to be quantified. In contrast, the mind is deemed to occupy no space (although it does occupy time) and, therefore, cannot be quantified by scientific means. That the nature of the mind manifests as distinct from the extended physical is evident in the difficulty of posing questions such as: What is the weight of blue? How far is it from Marxism to Capitalism? Where is love? The manifest mental

17 The brevity of this section is noted, and the current thesis acknowledges the complexity of Descartes' epistemology (e.g., see Newman, 2019) but length of the thesis is a constraint.

image, therefore, appears unmeasurable by standards of scientific measurement. As will be demonstrated under the discussion of physicalism in section 2.5, this is a category mistake in confusing things (substrates/entities) for how things hang together (what substrates do/processes).

If the mind is of distinct and separate substance to the physical, how is it that they interact (have a causal relation between the two substances)? Descartes acknowledged how the problem of mental causation forms a major objection to the mind–body problem for interactional dualism. In Cartesian dualism, the mind decides to perform an action (e.g., leave the chair for a glass of water) and the body responds to this and performs the action. The mind, therefore, causes physical effects and *vice versa*. For Descartes, this interaction was performed through the pineal gland that has no supporting evidence in modern neuroscience.

What neuroscience has demonstrated is that each physical effect (e.g., drinking a glass of water) has physical antecedents (e.g., neural activity that leads to the effect). These physical antecedents are acknowledged by both dualists and physicalists. The current thesis is unaware of any known mental event that has not occurred without a physical correlate (all neuroscientific investigations of mental features have shown a change in physical neural activity). Even mental events that may be considered purely mental, such as dreaming (there is no interaction with the environment), still has correlating brain activity (Dresler, *et al*., 2012). This is not to assert that these correlates are clearly understood, but simply to state that whenever there is a mental event a physical activity (substrate processes) occurs simultaneously. Therefore, if an attempt to explain an effect has one accepted antecedent (physical) and another is posited (here meaning the non-physical mental antecedent), certain problems arise. First, the law of parsimony may be invoked and query what the need for a second antecedent is, where one will do. This relates to the problem of over-determination, which asserts that it is unlikely that two causal antecedents would be active simultaneously for each effect (e.g., both the non-physical dualist mind and the physical neural activity for each physical effect of a person's life).

Furthermore, Descartes' (1637/1968, see also Lowe, 2006) argued that the mind appears unified and indivisible (indivisibility argument), whereas the body may be divided (e.g., removal of a foot has no impact on the unified nature of the mind). Hume (1777/1975), from within an empiricist epistemology, came to the opposite conclusion in finding the mind to be a diversity (a bundle of perceptions) and concluded that there is no unified self. Within clinical psychology, the divisibility of the mind is further established in many symptoms and diagnosis (American Psychiatric Association, 2013). For example, individuals with dissociative identity disorder experience "alterations of sense of self (e.g., attitudes, preferences, and feeling like one's body or actions are

not one's own)" (p. 292), schizophrenics may experience delusions, where "... one's body or actions are being acted on or manipulated by some outside force (delusions of control)" (American Psychiatric Association, 2013 p 87). Furthermore, the mind appears divisible in common language usage, such as when under emotional distress a person may state, "I'm falling apart/I need to pull myself together/I've lost a part of me". In essence, the mind appears indivisible sometimes (emphasised by substance dualism) and divisible in others (emphasised by empirical research). If the mind is a unified, non-divisible thing/entity, then these experiences are problematic (how can that which is non-divisible divide?), however, if the self/mind is a description of diverse processes that at times interact more (appearing to be unified) or less (appearing divisible) cohesively, both experiences may be accounted for (see Dennett's 1992 view of the self as analogous to a centre of gravity).

The final Cartesian argument turned to here is the argument from conceivability, where it is conceivable that the mind can exist without a body. The idea is that what is conceivable is possible, if it is possible that I can exist without a body, then my body cannot be an essential property of the mind, and, therefore, the mind is not physical. One way to respond to this line of argument is to query whether conceivability is a valid method of establishing possibility. Kim (2011) queries the validity of this line of reasoning by discussing Goldbach's conjecture, the assertion that "every even number greater than two is the sum of two primary numbers" (ibid., p. 39). At present it can be conceived that this assertion is both true and false, as both cannot be true; conceivability is a poor method for establishing metaphysical claims. Lowe (2004, 2006), a current substance dualist, also acknowledges the limitations of the conceivability argument. Conceivability and other forms of intuitive premise assertions are also open to alternative intuitions, which may contradict the initial argument. For example, I find it inconceivable that a mind can exist without a physical instantiation. If what is conceivable to one is inconceivable to another, then much rests on the assumptions brought to the discussion. Similar arguments could be made to affirm mind uploading, for example, is it conceivable that the mind could be instantiated across multiple platforms? If so, and we follow the logic of the above argument, mind uploading is feasible. However, this thesis has contested whether conceivability should be adopted when considering the metaphysics of mind.

For Descartes (1637/1968) and other substance dualists, the disparity between the mind and the body refers to the mind as a different kind of thing than the physical body, a different substance, a substrate independent of the physical substrates, and so on. As the thesis elucidates later (e.g., section 4.2), the mind is rather a process (what the substrate does) than an entity (what the substrate is), and this is the primary category mistake initially suggested by Ryle (1949). This

implies, therefore, that Descartes may have been confusing, to use Sellars' terms, the thing with how things hang together. Substance dualism has continued to develop since the time of Descartes (for example, see Loose, Menuge, & Moreland, 2018 for a collection of works in this regard).

2.4.2 Evaluation of substance dualism

Substance dualism has been critiqued from multiple perspectives, from behaviourists (e.g., Ryle (1949), identity theorists (e.g., Smart, 1959), property dualists (e.g., Chalmers, 1996), with these, and more, perspectives being developed further throughout the thesis. Within this section, the current thesis notes that substance dualism, as discussed above, fails in one of two ways. First, it has not found a solution to the problem of mental causation and over-determination. Second, it relies too heavily on intuitions based on the manifest image.

A simplification of the dualist argument may be presented as follows:

(1) The mind appears to have feature x.
(2) The body does not appear to have feature x.
(3) Therefore, the mind and body are separate entities.

Each of these points can be contested and particular intuitive assumptions are needed for substance dualism to be retained. A physicalist response may offer some of the following criticisms. In relation to 1) first, the appearance of feature x does not equate to feature x being real as it may be an illusion. Second, if the feature x is asserted to be real it is either based in mental substance or a physical substance. If a mental substance is asserted, the problem of mental causation and over-determination is retained. If a physical substance is retained, these problems may resolve (see section 2.5). This resolution would then raise other problems, such as how a physical feature can be understood as, or result in, a mental feature that is the subject of the rest of the chapter. The second premise 2) asserts that it appears that the body cannot perform what is identified as a feature in 1). For example, Taliaferro (2018) argues for the distinction between the mental and the physical by noting that a person ceases to exist on biological death yet the biology/physical (the corpse) remains. This is termed the corpse problem here[18], where the biological matter may exist yet the person/mind is said to not be present. As physical monists may assert that the person is a physical process (in which case, the disparity is accounted for, in that the mind is what the substrate does

18 The corpse problem is also related to the persistence problem (e.g., Shoemaker & Strawson, 1999; Olson, 2004), which is the subject of the following chapter (see section 3.5.).

and not what the substrate is), it is unclear why the corpse problem cannot be understood as when there is a cessation of *process* there is a cessation of person, which need not imply any substance distinction. An uncontested observation is that upon biological death there is a cessation of processes (e.g., the heart, lungs, neural activity, and so on, cease to function). The physical monist may infer from this cessation of function that the mind (which is deemed to also cease by both the physicalist and substance dualist) is, therefore, likely to be one such further process. The dualist is, therefore, making a category mistake by asserting entities where processes suffice.

A second confusion may arise within dualism in confusing entities for representations. If the mind is a representational system (to be discussed further in section 2.6.2.), it would be akin to confusing the map (the representation) with the territory (the referent). The features of a representation (e.g., the map being a piece of paper) is different from the territory (e.g., hills, vegetation, and so on). Therefore, it is acknowledged (by both dualists and physicalists) that the mind appears to have different features, but the dualist asserts entities beyond the physical, whereas the physicalist need not posit further entities. These distinctions between processes and entities are further developed throughout the thesis when various physicalist philosophies assert this in various ways. As both 1) and 2) in the above argument can thus be queried, then so can the conclusion 3). Note, that this does not necessarily critique the logical validity of the argument but queries the soundness of the conclusion if the premises may be queried as veridical.

This evaluation section ends with a thought experiment aimed to explore to what extent a mind may be deemed independent and non-extended (two of the assertions that are used to validate substance dualism). Suppose an alien space ship were to suddenly vaporise earth (call this the psychopathic alien experiment). Assume that all persons (e.g., astronauts) are on earth at the time. Through this act, all likely extendable physical properties and processes related to minds are removed from existence. Would the mental (e.g., the love for one's wife) remain once earth has been vaporised? If it is asserted that the mental would not remain, then minds are in some sense extendable, spatially located on earth, and dependent on the physical (if there is no physical earth including all physical substrates and their interactions there is said to be no mind). It is this type of dependence that is needed for the MUP, in which case, the mind needs a physical base and may include internal physical processes (e.g., neural mechanism) or external processes (e.g., behaviours). Alternatively, it may be asserted that the mental would be retained, and the mind is not spatially located at all (i.e., some kind of a soul substrate). In which case, the MUP would not be feasible.

2.5 Physicalism

As stated earlier, physicalism is the current dominant view within philosophy of mind. The physicalist view for the MUP is a negative hypothesis denying supernatural entities or intervention, while allowing any natural phenomena within the purview of human intervention and artifact creation. Physical monism, the view that the mind is constituted by a physical substance, is the alternative to substance dualism and views on physical substance may be categorised as those who hold to standard physical properties and those who assert that the mind is also constituted by further physical properties. The current thesis begins with views that are aligned with standard physical substrates and later incorporates views that assert substrates with further physical properties. Physicalism asserts that everything that is ontologically real is the result of physical entities and processes (however defined). As with all philosophical views, there is a plethora of arguments for and against physicalism, as well as within the physicalist stance, with the current thesis aiming to sketch and develop broad concepts within the field in relation to the MUP.

An argument from causation for a view based on a standard physical substrate may be presented as follows (adapted from Papineau, 2000):

(1) All physical effects are determined by physical antecedents (the completeness of physics).
(2) Mental events have physical effects (causal influence).
(3) Physical effects are not over-determined (no universal over-determination).
(4) Therefore, mental events are physical (conclusion).

Premise 1) refers to the premise that physics is complete. Completeness is not meant to imply that it is currently a fully resolved field of enquiry, but rather completeness refers to the assumption in approaching the domain of the physical. Note, that this is not a physicalist (that everything including the mental is physical) assumption in and of itself, because it says nothing about non-physical effects and antecedents. Papineau (2000) presents two arguments as to why this is the consensus among scientists who work within the physical domain.

First is the argument from fundamental forces that Papineau (ibid.) likens to the historic debate over vitalism and the current debate over the mental. In vitalism, it was previously assumed that living entities had a causal power over and above the fundamental physical forces developed in physics (related to conservation of energy). This strategy of assuming a something over and above (a separate life force, or *elan vital*) factor occurred historically in vitalism (since rejected)

and occurs currently within philosophy of mind in relation to the mental (still present). In relation to vitalism, Papineau identifies Helmholtz and his colleagues as those who championed the reductionist view that physical forces are all that it is needed, and which provided a more elegant solution than expanding theories to include alternative forces. However, elegance (or parsimony) was not sufficient in itself to sway the scientific community and, therefore, the second empirical argument from physiology was needed.

The argument is that the growing body of physiological experiments and theories that have been explored and developed up to date have found no evidence of any vital force, and all physical effects that have been observed may be explained by fundamental forces. It can, therefore, be inferred that the same will occur in relation to the mental as neuroscientific investigation increases our understanding of the mind and no non-physical domain (over and above the physical) has yet been discovered. This does not rule out the possibility of some future experiment being conducted that would call this into question, but simply underlines that ongoing and detailed analysis from a multitude of research has found no such evidence yet, making the discovery less and less likely. Both of these arguments may be contested, however, as the scientific momentum and evidence continues to mount in its favour; premise 1) is at least well founded and, at best, should be outright accepted. The empirical trajectory, therefore, favours premise 1).

Premise 2) allows for mental events to form part of the causal process and, therefore, avoids epiphenomenalism. The mind may well be an epiphenomenal physical effect but this possibility is excluded at present in an attempt to address the intuition of the interactionalist dualists that the mind has causal influence. The causal argument for physicalism here, therefore, accepts the desire to avoid epiphenomenalism and produces one logical argument for physicalism that retains the mind as a causal feature. Within a physicalist epiphenomenal/illusionist, stance 2) may be removed and 4) could be adapted to "Therefore, mental events are epiphenomenal/illusion". However, let it be assumed for now that the mental is causal[19].

Premise 3) relates to the concept of over-determination. If there are two necessary and sufficient causes for the same event, it can be said to be causally over-determined. An example of a man who dies by being shot and struck by lightning simultaneously may be introduced (Papineau, 2002). In this case, both antecedents (being shot and being struck by lightning) would be necessary

19 It may be inferred from evolution that the mind is likely causal because most evolved features (here including the mind) serve adaptive functions and play causal roles in reproduction and survival (Dawkins, 2011); alternatively, non-adaptive features, such as Gould's (1997) spandrels, are less likely.

and sufficient conditions in their own rights (independent of each other) to produce the same result (the death of the man). What is being referred to in the over-determination instance, is that a particular event has two complete independent yet equally effective causes simultaneously. Although it may be possible that a world may exist, where there is continuous parallel causes, this is deemed by the current thesis to be improbable and counter to the scientific method (which operates off the law of parsimony in relation to cause and effect). If over-determination is denied and physical antecedents accepted as causal (as is the mind), then the mind must be a physical phenomenon.

The problem for physicalism is, therefore, not of causation, indeed, as presented above, causation acts as an argument in favour of physicalism, but rather in the apparent contrast between the physical (the scientific image) and the mental (the manifest image). How can concepts such as qualia, what-it-is-like, the authority of the first person, and so on, be understood if everything is ultimately physical? First, the thesis turns to views that assume a standard physical substance/substrate with variation in emphasis on external and internal processes.

2.5.1 External processes (standard physical substrate)

In this section, the focus is on views that maintain that the mind is based in a standard physical substrate but is identified with or defined by external processes of that substrate. This means that the substance (the thing and the parts of the thing), where the mind is instantiated is constituted by standard physical properties (e.g., mass, energy, atoms, molecules). The relation to the multiple realizable physicalism argument is that if it is the external processes that are (in some sense) the mind, then replication of these external processes would be necessary and sufficient conditions for the MUP.

2.5.1.1 Behaviourism

Behaviourism asserts that the mind is the physical processes of the organism (entity/substrate whole) in its environment[20] (externally observable processes). This view will be developed throughout the section and it is noted at the outset that external processes relate to Sellar's things/entities and how they hang together and the category mistake of asserting the mind is a thing (entity) when it is actually related to how things hang together (process). The emphasis for the behaviourist is not to focus on the inner workings of the mind, but rather to see the mind as what

20 The current thesis primarily addresses the MUP within the context of a standard environment (the world we live in). Alternative environments, such as virtual environments, could be considered, but this would expand the MUP project to include world emulation alongside mind/brain emulation, which would increase the engineering complexity. The current thesis, therefore, opts to focus on standard environments.

the entity (the body) does in its environment. This section addresses two forms of behaviourism; namely, the analytic/logical behaviourists and the methodological behaviourists.

The first type of behaviourism that is presented in the current thesis is that of logical/analytic behaviourism that Hempel (1949/1980) directly associated with physicalism and the need to understand psychology within empirical science. Ryle (1949) and Wittgenstein (1953) are discussed in this section and the reader is encouraged to explore Carnap (1991) and Quine (1995, 2013) for further logical/analytic behaviourists' perspective.

Ryle (1949) criticised modern philosophy of mind, which was at the time still largely influenced by Cartesian dualism, as being grounded on a "category mistake" leading to "the dogma of the ghost in the machine" (ibid., p. 17). The mistake rests on categorising types inappropriately with the emphasis for Ryle, who developed his thoughts within the analytic tradition, on the usage of language. As discussed earlier (section 1.5.2), Ryle presented category mistakes, such as the foreigner who saw elements of the university (the library, the hall, etc.) but still asked where the university was (implying some entity over and above the elements). A further example by Ryle is an observer of cricket who could identify game play but could not find the illusive entity of team spirit. The mistake being that team spirit is not a thing to be seen but rather a behaviour – or sum of behaviours – to be observed. When Ryle (ibid.) turns to the subject of the mind, he sees the same type of mistake being made. To assert that the mind is an entity residing within body is the result of the category mistake that confuses behaviours (external processes) with substances (entities). For Ryle, we should, therefore, be cautious not to confuse the substrate itself as an entity with what the substrate does.

Ryle (ibid.) considers, among many 'mental categories', how intelligence is understood. For Ryle, the category of intelligence as a 'mental property' dissolves into incoherence if we remove the behaviour of acting within an environment to achieve goals. Without the observable processes there is no meaningful intelligence, but what of internal processes such as thoughts "in my head" (ibid., p. 36)? Ryle presents an example of a schoolboy performing a simple piece of arithmetic internally and a more complex one externally (through the use of a piece of paper as he works through the calculation). The behaviour of the internal processes is the same as the external processes and so both are behavioural. The distinction between private thought (that which I do in my head) and public thought (that which is observable to others) is not of a distinct category but of the location of the process.

Within Ryle's (1949) stance, it is implied that how the mind filters and orientates behaviours is through mental dispositions that are themselves the result of previous learned experience.

The substrate learns through conditioning as the substrate interacts with the environment and these conditioned responses strengthen or weaken the substrate's possible responses to various external stimuli (see below on methodological behaviourism). These dispositions are not themselves causes but are mental features that allow for behavioural outcomes. For example, the disposition of a glass to break can be termed brittleness. If the glass falls onto a hard surface, the glass will break as a result of the disposition (brittleness) in conjunction with the causes of gravity and the impact on the solid (an environmental disposition) floor. The precise definition of a disposition and its distinction from causes may be complex and dependent on further assumptions (such as the nature of causation). The distinction here indicates that causes occur from external stimuli, whereas mental dispositions are conditioned responses that are themselves the result of external stimuli and the organism's scope of possible responses.

Wittgenstein (1953) emphasised that the privacy (first person) of the mental is better understood as learned behaviour as the person interacts with public (third person) knowledge, such as language. In the beetle in the box thought experiment, Wittgenstein asserts that private pain is similar to persons each having a box and declaring to each other a beetle is in their box. No person is allowed to look into the box of the other and each person declares their knowledge of what a beetle is, founded on observations of their own beetles. In this scenario, there could be nothing in a box, an entity that continuously changes, an inanimate object, an actual beetle, and so on. Without access to public language there is no manner for meaning of "object and designation" (p. 100) to occur and the private knowledge becomes meaningless. In a similar manner (the experience of) pain, if limited to private sensations, and denied access to public knowledge such as pain language and learned behaviour, is equally meaningless. Smart (1959) commenting on Wittgenstein, states that Wittgenstein's view of pain, is "doing a sort of sophisticated wince" (ibid. p.141). It is, therefore, incoherent to speak of the mind without speaking of behaviours (including language actions).

Whereas behaviour is deemed to be an important factor for other theories (see below), it is the primary factor for the behaviourist. The nomological boundary at issue here is, therefore, that of the external processes (e.g., social psychology, behaviourist psychology, anthropology, and so on). The logical/analytic behaviourist makes the metaphysical claim that the nature of the mind is identical to the behaviours of the entity in its environment and may be illustrated as follows:

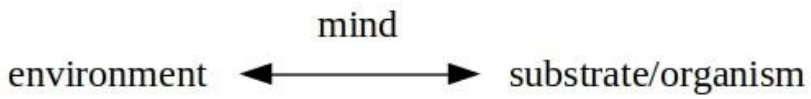

Figure 2-2: Behaviourism in terms of an external process

Whereas the logical behaviourists developed behaviourism from a philosophical stance, a similar approach was being developed within the field of psychology with similar themes. For example, the privacy of the mental is "distinguished by its limited accessibility but not, so far as we know, by any special structure or nature" (Skinner, 1953, p.257); inner 'mental' aspects, such as seeing the colour red, pains, itches, and so on, are seen as embedded within a language community, and there is no purely subjective mental ontology but rather an integration of behaviours and limitations of accessibility. Methodological behaviourism emphasises practical concerns as the preferred method for understanding (epistemology) the mind, rather than asserting a metaphysical claim. The methodological behaviourist may assert that observing behaviours has been the primary method for establishing the ontology of other minds throughout human history and the only defence against solipsism (the belief that the only mind we can be aware of is our own and that, therefore, our own mind is the only reality). If the method for verification of the ontological existence of a mind is purely introspective, then there can be no way of determining that others have minds, but if we can infer from behaviour (others act in a way similar to our own) that other minds exist, then methodological behaviourism becomes useful.

The methodological behaviourist emphasised the mind as a learned process as a physical entity interacts with its environment. Watson and Rayner (1920) showed how emotional responses can be learned; Pavlov (1927) worked with animals (most famous for his work with dogs) and showed how conditioning can occur through the pairing of stimuli (the dogs salivated to the sound of a bell that was previously paired with food presentation). Through these experiments, it became evident that a) behaviours need not have conscious mechanisms that may be termed dispositions; b) that aspects of the mind, such as learning, have biological mechanisms determined by environmental stimulus and response processes – this learning interaction between environment and entity that results in dispositions is commonly termed 'conditioning'; c) that introspection is limited in understanding processes of the mind; and d) observing behaviour is a credible and practical method in developing a theory of mind.

It is noted that when the substrate and the environment are observed, standard physical entities (each can be defined and described with standard physical measures) are being observed. If a mind is determined through observing behaviours and these behaviours occur within an environment, then to attribute a mind to a non-behaving entity is a mistake. For behaviourists (both logical and methodological), the idea of a non-behaving mind is similar to a non-flying flight, an unthought thought, and so on. Therefore, behaviour is a necessary condition for the metaphysic of the mind in the behaviourist context.

The idea of defining mental phenomena through observable behaviours was adopted (and adapted) by Turing (1950) in the field of developing computer reasoning (which was to become artificial intelligence). He presented the imitation game (now referred to as the Turing test) as a method to determine whether a machine could think in a similar fashion to a person. Therefore, thinking was taken to be the primary feature of the mind (intelligence). Consider a scenario where there are three persons; one man, one woman, and one of whom is an interrogator who has no direct contact with the other two persons. The objective for the interrogator is to determine who the man is and who the woman is. The objective of one of the two interrogated persons is to deceive the interrogator that they are the other interrogated person. Now consider that the deceiver is a machine. It is put forward, that if the interrogator cannot distinguish between a person and the machine, then the machine can be thought to have intelligence akin to human intelligence.

Although Turing (ibid.) acknowledges the possibility of biological technological machines, his main interest was that of developing a computational machine (Turing machine). Turing addresses some criticisms, such as the mathematical problems (here focusing on Gödels theorem and the Church-Turing hypothesis), that machines can't be conscious or creative, and so on. For Turing, these express the possible limitations of machines but he did not view them as sufficient refutations, as, according to him, the human 'machine' is equally limited and the limitations of future artifact machines are unknown, and, therefore, a test (irrespective of speculative limitations) can still be defended. The underlying assumption in this paper was that behaviour is the method (methodological behaviourism) best suited to determine the possibility of intelligence, as behaviour is the measure of whether the Turing test is passed or not.

2.5.1.2 Externalism

A further aspect of physicalism that relates to the mind and external processes is that of externalism. Externalism, in relation to mental content, is associated with intentional states, such as beliefs and desires, and the assertion that this intentional mental content is dependent on the relation to the environment (Lau and Deutsch, 2016). The current thesis addresses two thought experiments that infer the need for environmental interaction as necessary for meaning and belief, as found in the works of Putnam (1975) and Burge (1979/2002). It is noted that both Burge (2007) and Putnam (1965, 1980) elsewhere assert that the mind also includes internal processes and, thus, the emphasis of the current section is not on the entirety of their works in philosophy of mind but on their emphasis on external processes. If both internal and external processes are needed for a mind, then the external emphasis of the externalism may reflect a necessary but not sufficient condition of the mind.

Putnam (1975) introduces the thought experiment of a twin earth that is identical to Earth save in one manner; 'water' on twin earth is not H_2O but a complex chemical summarised as XYZ. Their 'water' (named 'twater' herewith) performs the same functions water on Earth does. It fills the lakes, is consumed to retain hydration, waters plants, it looks the same as water, it tastes the same, and so on. On this twin earth, there is a twin you who has lived an identical life to you in every way (except for his knowledge that water is XYZ, while yours is that water that is H_2O). If you were transported to twin earth, you may have the psychological state that results in you saying "that is water" as you point to twin earth's twater. Twin you may say the same sentence "that is water" with the same psychological state. Given that it can be questioned as to whether you and your twin are in the same psychological state, as the knowledge of the chemical compound of 'water/twater' is part of the psychological state, Putnam (ibid.) asks the reader to imagine the same scenario but in 1750 (prior to chemical reduction of water), where your ancestor (and twin earth ancestor) both used the term 'water' with exactly the same psychological (internal) state, yet they were referring to two different entities (H_2O and XYZ). The key point is that intended meaning is not purely dependent on internal psychological states but dependent on the current theories and norms of a given sociolinguistic context. In Putnam's words "cut the pie any way you like, 'meanings' just ain't in the head!" (ibid, p. 144).

A further point raised by Putnam (ibid.) that bears mentioning is in relation to indexicality or token-reflexivity, where language is understood from the perspective of the speaker (or community of speakers) and extension is varied from different contexts or tokens. For example, if a person states, "I have a headache" there is seldom confusion as to whether the term "I" refers to the hearer or the speaker. Although the hearer has used and understands the term "I" to refer to themselves, it is understood that when another says "I" they are referring to themselves (it has two extensions in this context, one relating to the speaker and the other the hearer). This applies to many phrases such as "now", "here", "that", "we", and so on. In relation to twin earth's twater, it could be said that the term 'water' is indexical as is our use of the same term. In this sense, meaning is not purely internal but indexical to the community of speakers.

Burge (1979/2002) introduces a similar counter-factual thought experiment in relation to mental content and external processes. Consider a person who has many beliefs about 'arthritis', including that he has had arthritis for years, that his wrist pain is more painful than his ankle pain, that stiffening joints is a symptom of arthritis, and so on. In addition to these beliefs, he further falsely believes that he has recently developed arthritis in the thigh. He reports this to his doctor who corrects him that arthritis is specifically joint inflammation and that the symptom of pain in

his thigh cannot be arthritis. Now consider the counter-factual thought experiment, where all these beliefs leading up to the doctor's consult remain the same. However, in the counter-factual scenario, the doctor (in alignment with the medical field of the counter-factual world) asserts that the belief of arthritis in the thigh is true because arthritis in this counter-factual world includes other rheumatoid symptoms. In the original scenario, the original belief "I have arthritis in my thigh" is false, whereas in the counter-factual scenario the same belief is true. As all beliefs and experiences of the individual remain constant, the only distinction between a false belief and a true belief is within the external beliefs of the external medical field. Therefore, beliefs depend on the external context and are not based purely on internal processes (for more on counter-factual thinking see Lewis, 1986 possible worlds arguments).

Both these externalist arguments indicate that the mind (here referring to aspects of the mind, such as meanings and beliefs) is part of a dynamic system that interacts with and draws from the environment (i.e., meanings and the veracity of beliefs are determined by their interaction with the context). Consider language development, which is closely associated with meaning and belief. Without the appropriate environment, a child does not acquire language (e.g., in the case of feral children – see Candland, 1993) and the particular language acquired is dependent on environment (e.g., a child raised in France learns French, in England it learns English, and so on). However, not all necessary processes can be said to be external, some are also internal (e.g., the brain). For example, consider damage to Wernicke's area of the brain, where the patient's ability to understand language is disrupted (Wernicke's aphasia). The inability of an internal process (neural activity in this example) impacts on meaning also. It is, therefore, likely that both external and internal processes are necessary for language (and, by implication, other aspects of mind).

Within the MUP, as presented here, the environment remains constant and it is, therefore, assumed that external processes (however defined) between substrate and the environment will be retained. Therefore, if the internal physical processes are maintained (within the new substrate) along with the external physical processes with the environment, it should not alter the nature of mind from an externalist/behaviourist perspective.

2.5.1.3 The extended mind

Clark and Chalmers (1998) developed the concept of an extended mind, which integrates aspects of external and internal processes. The theory is presented within a computational theory of mind

perspective and focuses on cognitions as information processing (see section 2.6.2)[21]. They present various cognitive activities that integrate external processes, such as performing calculations with a pen and paper (similar to Ryle's student performing arithmetic), the use of scrabble pieces in word constructions, and so on. "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 recognising it as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. Cognitive processes ain't (all) in the head!" (ibid., p. 8). If the cognitive system can be seen as a coupling of both internal and external processes, then the mind is deemed to be extended. This coupling needs to be reliable and it is asserted by the authors that reliable coupling occurs primarily in the brain but may also occur external to the body (as well as within the greater system of the body itself).

Clark and Chalmers (ibid.) present an example where two persons (Inga and Otto) wish to visit a museum. Inga remembers the street address of the museum and, thus, retains a belief about its location as she navigates to there. Otto, who suffers from Alzheimer's, uses an external notepad to record the street address and refers to it constantly as he makes his way to the location. The idea advanced here is that the belief (where the museum is) serves the same function on the external notepad for Otto as in the internal memory of Inga, implying that the mind may extend to the external environment.

The notion of the extended mind is distinguished from externalism by the authors, in that the external is largely passive within traditional externalism, whereas in the extended mind theory the external is active. Clark and Chalmers (1998) use an example of a fish as a swimming entity, where the process of swimming includes the water's kinetic energy (currents, eddies, and so on) interact to result in the process of swimming. If the fish were out of the water, it could not be said to swim and so the external environment plays an active role in the swimming process. There is, therefore, a necessary interaction between the external environment and the entity to achieve the desired process. However, the analogy may also serve to illustrate how internal processes (here internal is taken to mean bodily processes) should be the primary areas of interest as swimming is not possible without a swimming entity (i.e., water without a fish cannot be said to swim). In the examples presented by Clark and Chalmers (1998), it can be noted that areas that are considered to

21 It should be noted that although Chalmers co-authored this paper, there is a qualifier of "authors are listed in order of degree of belief in the central thesis" (1998, p.7) that indicates that this is more Clark's thesis than Chalmers. Chalmers' views will be presented with further details in section 2.7.1.

be extended are areas that are under the control of the physical entity (a pen and paper, the scrabble pieces), whereas other aspects of the environment that fall outside of the control system (for example, another person) are not considered as extensions of the mind. Therefore, for extended mind theorists, the idea that a mind is a control system is a necessary condition.

In the extended mind context, as with other representation systems, the plausibility of mind uploading is affirmed. "In the distant future we may be able to plug various modules into our brain to help us out; a module for extra short-term memory when we need it, for example. When a module is plugged in, the processes involving it are just as cognitive as if they had been there all along" (ibid., p. 11). As the nature of the mind, in this view, is an information process system (see section 2.6.2) that interacts with an environment (and, therefore, is extended), it can be said that an alternative system that performs these same processes would be the same control system (mind).

2.5.1.4 Evaluation of external processes

Kim (2011) identifies "three main players on the scene in the discussion on mentality: mind, brain, and behaviour" (p. 87). Behaviourism has emphasised that to understand and explain the mind it is necessary to understand behaviours of entities in their environment and, as such, that the mind cannot be divorced from these processes. Where behaviourism differs from the other physicalist theories is in the lack of emphasis on the internal processes that lead to the behaviour. The growth of neuroscience and the development of neural imaging have, however, allowed the scientific community to peer inside the black box of the brain and observe diverse neural activity in relation to external behaviours. Although the knowledge is far from complete, it would be unwise to ignore the growing body of data that explains aspects of the mind as internal processes. For example, through lesion studies, brain areas are linked to behavioural change (e.g., frontal lobe damage leading to frontal lobe disinhibition) or chemical processes, such as serotonin re-uptake being linked to feelings of depression. Furthermore, what of pathology such as locked-in syndrome (or paralysis), where the person is unable to externally behave (Jordan, Elsbernd, Sladky, 2019) yet has internal processes? The introduction of recent brain–machine interfaces (Milekovic, *et al*., 2018) has allowed these patients to communicate with the external world indicating that a mind was there all along despite the inability to behave outwardly. Therefore, it appears feasible to interpret that the mind is not exclusively behaviouristic in nature, although it is impacted on and developed through

external interactions. In short, there is overwhelming evidence that internal processes are significant contributors to behaviours[22]. Therefore, both internal and external processes (however defined) are likely to be necessary and sufficient conditions for the MUP.

Behaviourism brought into doubt the Cartesian view that the mind is a separate substance from the body and allowed the possibility of understanding the mind as a process to gain ground (the view that the mind is not what the substrate is but what the substrate does). Some of the distinctions between the mental and the physical lose their impact when introducing the notion of physical processes. For example, a question such as, "How far is it from Marxism to Capitalism?" is similar to asking, "How far is it from good team spirit to poor team spirit?". That both are nonsensical is not because they concern different substances but that a category mistake between a process and an entity has been made. This distinction between entity and process, as will be demonstrated throughout the thesis, offers an explanation (at least in part) of the manifest appearance of disparate features between the mind and body.

That one physical substrate can perform a process also increases the likelihood that another substrate could perform the same process (what is possible for one physical system can, in principle, all things being equal, be possible for another). If this is the case, then the process can be said to be multiply realizable and, as the emphasis is on a physical substrate performing said processes, such a view would then fall under the umbrella of multiple realizable physicalism.

Within the current thesis, external processes have been categorised within views based on a standard physical substrate. However, the same conclusion in relation to the MUP (emulate the external processes and you emulate the mind), irrespective of type of physical substrate (standard physical or further physical) could be asserted if identifying the mind with the external processes is maintained. Here a simplified argument of the external processes is presented as:

(1) The mind is defined by or identified with observable external processes (whole or in part) (the metaphysical claim).
(2) Observable external processes are observations of physical entities and their interactions (the epistemological claim).
(3) Therefore, any physical substrate that can perform the same observable external processes will be the same mind.

---

22 The shift in moving from behaviours to include cognitions is also evident in therapeutic psychology, where the current "gold standard" is Cognitive Behavioural Therapy (David, Cristea, and Hofmann, 2018), which includes private (internal) thoughts.

The first premise is the argument of behaviourism (which identifies the whole mind with observable external processes or behaviours) and externalism (which identifies part of the mind with observable external processes). This premise could be denied by those who assert that the mind is rather identified with internal processes or some further physical properties.

The second premise is a reformulation of methodological behaviourism and asserts that the current method for observing these external processes (including speech behaviours) is through observing physical processes. Each nomological domain (e.g., social psychology, behaviourist psychology, and so on) that is identified with the mind from an external processing perspective observes physical entities and their interactions. For example, speech is observed through physical sound waves and interaction with physical eardrums, and so on. This is not to say that all external processes are caused by standard physical processes, but that whatever the cause, what is observed are standard physical entities and their interactions. Other views (e.g., property dualists, functional emergentists) may argue that what is required for the behavioural outcome is an internal or further physical causal process. However, what is being asserted in this second claim is that from the epistemological perspective, the method (hence methodological behaviourism) to determine the existence of the mind would still be the same (i.e., the observation of behaviour of standard physical substrates and their interactions) irrespective of the cause of said observable external behaviours.

The conclusion in the context of the mind as external processes is that any physical substrate that performs the same process is the same mind (based on the first premise of this argument stating the nature of the mind). The substrate contingently needs to be physical because these are the only substrates that we are able to observe (we can only observe physical processes). Therefore, to emulate a particular mind (in whole or in part) would be to emulate the particular external processes, and persons/minds are, therefore, distinguishable, based on variation of patterns of external processes. Because the current objects of observation of external processes are physical, a physical artifact (so long as it performs the same external processes) would be the same mind.

External processes of standard physical substrates offers, therefore, a necessary criterion for the mind to exist, although it is may be queried as a sufficient condition (as internal and further physical properties may be needed to cause the behaviour contributing to identifying the mind at any given time).

2.5.2 Internal processes (standard physical substrate)

Whereas a discussion of external processes emphasises the external physical processes that occur between the substrate (entity) and the environment (further entities), those who emphasise internal processes focus on processes that occur inside the substrate (substrate components). That there are correlations between mental phenomena (e.g., cognitions and emotions) and neural activity is not contested within modern neuroscience. The question that arises is what are we to make of these correlations and are they indicators of causation?

Kim (2011) explores the concept of correlation and causation, noting that when a systematic correlation is noted, a need to explain the correlation in causal terms is a common feature of scientific enquiry. Events can be causally related, for example, a lake freezes when the temperature drops below a certain degree. Events may have a further common cause, for example, when gas is kept in a rigid container temperature and pressure co-vary as both depend on kinetic energy of molecules. In this sense, temperature is not said to cause pressure, as they are both collateral effects of the common cause (molecular kinetic energy). Events may be identical, for example, lightning occurs when there is electric discharge because lightning *is* electric discharge (they are identical events). For Smart (1959), this indicates that the term 'correlate' may be misleading, as it implies that there are two separate entities (whether two different properties or substances) that need to be correlated, namely the mental and the physical. The current thesis, therefore, accepts the term 'correlate' with caution and uses it as a loose descriptive association between the manifest image and the scientific image.

2.5.2.1 Identity theorists

Identity theory focuses on the central nervous system and the brain (primarily) as the appropriate physical area of investigation in relation to the study of the mind and is also termed neurocentricism (Place, 1956; Armstrong, 1968; Goldman and de Vignemont, 2009.). As stated earlier, there is overwhelming evidence that supports this hypothesis, such as the impact of loss of neural function leading to direct loss of function in the mind. The neurocentricist does not assert that the environment and the greater body (e.g., stomach) have no impact on the mind. Rather, the neurocentricist claim is that the primary system of processing and interpreting information that is associated with the term mind, is that of the brain as it connects to the external world through the nervous system. As the brain, in this view, is seen as a processing system, it requires information to process and acknowledges that external stimuli form part of that necessary system integration.

The current thesis introduces here the thought experiment of the unlucky soldier to intuitively argue for neurocentricism (here emphasising the brain as the primary physical organ correlated with the mind). Consider a soldier who is captured and placed in solitary confinement. Will the loss of social interaction between the substrate and the social environment, result in a loss of mind? Now consider that he is freed but is later in an accident and loses all his limbs, and then later his stomach, his liver, his eyes, his ears, and so on. If a person can retain their mind despite the loss (e.g., of social interaction or body parts), then these should not be the primary physical system identified with a mind.

A gruesome story, yet each loss has a counter example in medical history where a person retained their mind while suffering loss until we encounter loss of the brain. There is no known account of a loss of brain with the retention of the mind. It could be argued that loss of brain indicates loss of life so let us consider a scenario where the brain is removed but the rest of the body is kept functioning, i.e., the lungs keep breathing, the heart keeps pumping blood, and so on. In this scenario, would the body be said to have a mind? To this we could explore medical history when a person is considered 'brain dead'. What is observed in these instances is that there is no indication of a mind (by any meaningful definition). In essence, a person can change environment and lose any body part and retain a mind yet cannot lose the brain. Therefore, the brain is likely the essential physical entity of the mind.

Whereas the behaviourist made the identity claim between the mind and behaviour (external processes), the identity theorists make the claim that the mind is identical to brain processes (internal processes). Smart (1959, 2017) describes the identity theory as a "species of physicalism" (Smart, 2017) that denies strong emergence (and is, therefore, reductionist) and is an ontological but not translational claim (see also Feigl, 1958/2002; Place 1956). It is noted at the outset, that the assertion here is not that the physical substrate (e.g., the brain) is the mind but rather the mind is identified with the processes of the brain (what the brain does). For example, if pain is identified as C-fibre firing (or any other such internal biological process), then it is noted that it is not identified with inactive C-fibres but with the process (firing) of the substrate (C-fibre). In this view, a corpse has C-fibres but has no pain, because there is no stimulating of these fibres. Thus, it is asserted in identity theory that pain may be identified with the processes (C-fibre firing).

A focus for these identity theorists, who primarily work within analytic philosophy, is how the use of language has caused confusion and may be better understood to identify the mind with the physical processes of the brain. First, language in terms of folk psychology about the mind is not translatable into the language of the physical sciences. For example, the term 'sensation' is not

translatable to terms such as molecules and dopamine (Place, 1956, Smart, 1959). The identity theory does not claim that a mind *is* a brain, which lead to problems statements such as "a mind weighs 1.5 kilograms". This would make a category mistake of confusing entities with processes. Place (1956) asserts that the argument against the mind (consciousness), being identical to brain processes, rests on the failure to distinguish between the 'is' of definition and the 'is' of composition. Within the 'is' of definition, it makes sense to add the addendum 'and nothing else' and he presents examples such as 'red is a colour', 'a square is an equilateral rectangle' and so on. There is a similarity between this type of 'is' and a logically necessary statement. However, this is contrasted with the 'is' of composition, where the addendum 'and nothing else' is illogical as well as relying on contingent factors. For example, it is nonsensical to state that 'he is old and nothing else' or to say that 'the table is made of wood and nothing else'. The 'is' of composition gives us useful information and may be contingently true (e.g., the table may be made of wood if contingently verified) yet it serves a distinct linguistic function from the 'is' of definition. The 'is' of composition is the function that Place refers to in the statement 'consciousness is brain processes'. In this sense, that consciousness (or a specific mind) 'is' a neural activity (the identity) allows for the possibility that consciousness (or a specific mind) may also be constituted ('is') by another substrate kind. Just as the table may be made of wood, it may be made from metal; and so, the mind which 'is' neural activity may also ('is') be constituted by another substrate and is, therefore, multiple realizable.

Furthermore, Place (ibid.) introduced the notion of a 'phenomenological fallacy', where the description of an experience (sound, taste, pain, and so on) is mistaken for actual properties, as they appear in some form of internal theatre. For Place, consciousness (here linked to linguistic descriptions of experiences) is preceded by discriminatory systems (it can be assumed that animals and infants have discriminatory capacity without linguistic consciousness under discussion) and it is only once we have obtained linguistic capacity (the ability to describe the outside world) that consciousness emerges. These experiences of consciousness are, therefore, in this view, higher-level representations (see section 2.6.2) and the phenomenological fallacy lies in conflating representations with objects of representation (referents).

Smart (1959) offers similar direction in relation to representations. For example, in relation to sensations and brain processes, Smart associates the sensation to the report of an after-image (in his example a roundish colour image that is orange toward the centre and yellow towards the edges). Peering inside the brain these colours are not observable, so in what sense do these colour sensations (sense-datum) exist? To what is the statement "I see in my mind an orange-yellow blurry circle" reporting? For Smart (ibid.), it is reporting a neural process that is an aggregate of past

processes. The brain processes can be seen as heuristics of external stimuli (generalised simplification of external stimuli) as well as representations (that which can be retained when the stimuli is no longer present). The report is akin to the statement, "there is something going on which is like what is going on when …" (ibid, p. 287) I am alert and presented with similar stimuli. The brain-process is not the after-image (which has no ontological physical existence) but the sense-datum (the experience of the after image) is a brain process.

Furthermore, identity theorists (Place, 1956; Smart, 1959, 2017) do not necessary hold that identity between mind and neural processes are necessary *a priori* truths (see Lewis 1966; Armstrong, 1968 in relation to *a priori* identity theory), but rather assert they are contingent truths established through empirical research. Smart (1959) presents the argument that the 'evening star' and the 'morning star' are both contingent on the existence of Venus, with the evening and morning descriptions referring to the same entity (the planet Venus) but observed at different times related to the earth's rotation. The term 'evening star' is not of identical meaning to the term 'morning star' and, therefore, they cannot be reduced to the same linguistic meaning. However, they do refer to the same entity. A further example of Smart's (ibid.) and Place (1956) is that of lightning and electrical discharge, where it can be said that lightning *is* electrical discharge (it is identical) yet descriptions of lightning are not the same as descriptions of electrical discharge.

Smart (2017), in relation to the type–token distinctions, asserts that it may not be "an all or nothing affair" and that there may be "no ontological difference between identity theory [commonly, type identity theory] and functionalism [commonly, token identity theory]". Consider Table 2-1:

Time	**Physical**	**Type identity physicalism**	**Token identity physicalism**
Substrate A (the current biological body)			
t_1	P_1	M_1	M_1
t_2	P_2	M_2	M_1
t_n	P_n	M_n	M_1
		$M_1 – M_n$ may allow for the same mind to instantiate the same pain	M_1 refers to the same mental state
Substrate B (the MUP artifact)			
t_{n+1}	P_{n+1}	M_{n+1}	M_1

Table 2-1: Type-token identity distinctions in the mind–body problem

First, consider the token identity physicalist option in the far right column. In this formulation, multiple events of physicalist states (P_1–P_n) may be designated as referring to a single mental state (M_1). For example, pain may be categorised as a singular state (being in a state of pain), while acknowledging multiple physical instantiations (the different physical tokens of $P_1 – P_n$) produce that same mental state of pain (M_1). It is acknowledged that there are variations within the substrate yet these variations do not change the identity claim that there is the same mental state of pain. Furthermore, substrate A is continuously changing (e.g., aging, neural plasticity, ongoing replacement of atoms and molecules), yet the pain (my pain, your pain) retains the same mental state (M_1). For the MUP, the question is whether an alternative substrate (Substrate B) may instantiate the same pain (the pain of the particular person). If substrate A may change and yet retain the same mental identity of a particular mind's pain, it stands to reason that it is possible that a further change (P_{n+1}) may instantiate the same pain (M_1) and then the MUP would be feasible because M_1 occurs across substrates.

Second, consider the type identity physicalist column, where for each unique physical state there is a unique mental state, such that P_1 is the same identity type as M_1. Some (e.g., Bitchel & Mundale, 1999; Barrett; 2013) consider that the variation of physical activities (P_1–P_n) undermines type physicalism and multiple realizability, however, this has been contested by others (e.g., Smart, 2017; Jackson, *et al*., 1982; Akand, 2018). Assuming that there is indeed variation of physical activities at different times such that pain at t_1 is a unique physical state (P_1) that is also a unique mental state (M_1). This pain type is distinct from the pain at t_2 which has the unique physical state (P_2) and the unique mental state (M_2).

If it is accepted that M_1 and M_2 are different types, can it still be claimed that both M_1 and M_2 are the pain of the same mind? In essence, if it is acknowledged that a numeric mental state is identified with a numeric physical state, does this negate the idea that the same mind's pain (despite variation of physical states) can continue over time? Consider that if physical variations of P_1 to P_n are accepted as identified with M_1–M_n, and that these variations are nevertheless the pain of the same mind (your pain), then it follows that the same mind's pain is multiple realizable at different times, with different physical variations. Pain may be too broad a term, so consider a more specific pain, such as stubbing ones toe. The same reasoning of physical variations, mental variations, yet retention of the same mind's pain can still be upheld (i.e., it would be acceptable to state that I had the same pain of stubbing my toe on alternative nights). The same mind's pain, so defined, is, therefore, a further description (a further type according to the type sortal spectrum) that transcends across multiple times (different experiences of pain) and this more abstract type description (my

pain of stubbing my toe) is, therefore, multiple realizable (i.e., M_1–M_n may be a further abstract type, such as the pain of stubbing one's toe despite it being constituted by different type states of M_1–M_n).

These P_1 to P_n and M_1 to M_n all occur within the substrate A (the same numeric body) and, therefore, the question may be raised whether this pain of the same mind may occur in an alternative substrate (substrate B). If the same mind's pain can occur with physical variation in the substrate A, it is feasible that this same mind's pain may occur in substrate B if sufficient continuity is upheld. If M_1–M_n within the substrate A is the pain of the same mind, then it is possible that a similar enough variation in substrate B (P_{n+1}) may instantiate a continuity of the same mind's pain. However, if it is claimed that only a particular substrate leads to particular mental phenomenon, then this is the stance of the flat realizer (discussed earlier in section 2.2.1) and faces the problem of how this substrate may have unique properties and processes (the uniqueness problem mentioned in section 2.2.1). If a type identity theorist were to maintain that only a particular substrate can instantiate a particular mind (that continuity from M_1–M_n is feasible but not M_{n+1}), then this would negate the MUP as it is akin to the biological solution of the persistence problem addressed in chapter 3.

For both type physicalism and token physicalism (as defined here), the mental is caused by, or subsumed under, the standard physical processes and both have the physical categories of P_1–P_n as designators. As both type physicalism and token physicalism have the same standard physical substrate processes (standard physical properties and processes such as neurons firing), it can be said that replication of these similar processes (P_1–P_{n+1}) may both instantiate the same mental phenomenon (however defined). If the identity claim of either type or token identity theorist were to be that the mind is identified with the substrate (as opposed to what the substrate does), then the MUP would not be feasible (i.e., the mind may only be instantiated in a particular substrate). However, if it the identity claim is that it is the process that is the mind, it is possible (e.g., Place, 1956; Smart, 1959, 2017), in principle, that an alternative substrate could continue this process. This retains the type identity claim that it is the substrate that is performing the process (what the substrate does) but denies that the same particular substrate (i.e., the 'is' of composition) needs to perform this process.

Lewis (1972), similar to Feigel (1958/2020), states the basic argument for his identity theory as such:

(1) Mental state M = the occupant of causal role R (by definition of M).

(2) Neural state N = the occupant of causal role R (by the physiological theory).

(3) Therefore, mental state M = neural state N (by transitivity of =)" (ibid, p. 249).

Lewis (1972) is cautious about whether mental states, as described in common language, necessarily refer to ontological entities but accepts that the collective human experience (the causal theory of the mental such as in the DBA model) is a myth and asserts that "I adopt the working hypothesis that it is a good myth" (ibid., p. 257). He acknowledges that theories may adapt and terminology may change as theories are constructed and evolve. What is determined to be real are the entities within the theory that fit a causal explanation and are validated through reasoned investigations. The general process is then to examine each of these mental states (one assumes as expressed through self-report) and correlate them to physical activities and, therefore, the starting point for his version of the identity theory is the acceptance of the mental as real. If a consistent pattern is noted, then the two (mental and physical) may be identified as the same entity. However, the myth may be queried, which is the stance taken by the eliminativists in the following section (section 2.5.2.2.). Lewis's ideas of how a type identity theorist may still retain psychological continuity (how the same mind may be retained despite variation in physical substrate) is further developed in section 3.5 of chapter 3.

Identity theory is a physicalist metaphysical stance and may be formulated as Ll (lower level) = Hl (higher level), in that the psychological (Hl) is, in some sense, a description of the neural activity (Ll). It is, therefore, a view that emphasises the standard physical substrate and its processes. But what can be said in relation to multiple realizability? Smart (2017) asserts that many identity theorists are also functionalists in some sense, including Lewis, Armstrong, Place, and himself. Smart (ibid.) uses the example of the eye's function to see, although the term can have different physiological realizations, such as the distinction between the eyes of a dog and a fly. Furthermore, Smart explores desires and beliefs in relation to a robotic self-flying aircraft. In this example, the aircraft is programmed to fly between Melbourne and Sydney and in order to do this it would need a representational system (a map) that would function as the aircraft's beliefs about the world. It would also need a desire to stay on course such that any deviation (such as initiated through adverse weather) would engage corrective measures to course correct. For Smart (ibid.), what this indicates is that there are no metaphysical semantic or intentional problems and this difference between brain processes and robot processing would, therefore, be a matter of information complexity (engineering) and processing method rather than anything regarding a different metaphysical domain.

What this indicates is that not only is multiple realizability consistent with some versions of identity theory but that the computational mind may also be accepted within some versions of identity theory. Because identity theorists emphasise biological processes, the question arises as to whether non-biological processes (such as information processing that could occur in a synthetic computer) can achieve similar enough processes. If only biology can perform these processes then the MUP would need to include biological artifacts. If another substrate (such as silicon based artifacts) could produce the same processes, then a more synthetic artifact may be used. And so, identity theory may relate to both the A (artificial) and B (biological) of the ABC options in relation to mind uploading. Neither the emphasis on the biological (carbon based) nor the synthetic (silicon) artifacts negates the feasibility of mind uploading, but they both introduce potential empirical and engineering restrictions and directions for the project.

2.5.2.2 Eliminativism

Eliminativists accept that brain processes are that which the mind is (and so is a form of identity theory) but call into question whether our current anecdotal understanding of the mind is necessary or sufficient for a philosophy of mind (Feyerabend, 1963; Churchland, 1981; Churchland, 1986; Churchland & Churchland, 1997). The concept of mind, as understood by most persons in their everyday lives, can be referred to as folk psychology, where mental processes are described as a causal theory where mental events impact on behaviour (such as the DBA model). The question then arises as to whether it is a sufficient theory and whether it should be retained or eliminated (or at the least redefined) in the light of scientific theories such as those developed in neuroscience.

Feyerabend (1963) likens folk psychology to previous explanations of epilepsy as demon possession. The scientific knowledge of the physical decreased the need to define symptoms such as convulsion and loss of consciousness as demon possession, and a preferred scientific image (further validated through effective medical treatments) has been embraced. Churchland (1981) likened folk psychology to alchemy, which has also been abandoned in preference to the science of chemistry. Within this analogy it can be noted that many of the terms were retained (e.g., gold, silver, elements) but the causal explanation (e.g., that four fundamental spirits of mercury, sulphur, yellow arsenic, and sal ammoniac were causal and combinational to chemistry) was not.

Theory reduction is an important component of the eliminativist perspective (Churchland, 1986). Churchland (ibid.) explores many scientifically reduced theories from the past and notes that, in most instances, the previous theoretical concepts are not directly reduced to the new scientific concepts, but that commonly there is an abandonment (elimination), or at the least refinement, of concepts. For example, the caloric theory of heat (that heat is some form of self-propelling liquid

that moves from hotter to cooler entities) was eliminated and replaced by the kinetic theory (that heat is the kinetic energy due to molecular movement). Elimination is not the only option within theory interaction. For example, the electrical and magnetic theories of the past have been integrated (and significantly adapted) into the electro-magnetic theory of Maxwell (further refined by Heaviside – see Hampshire, 2018) which integrated Gauss' law, Faraday's law, and so on. In this instance, electrical theory and magnetic theory were not reduced in either direction but rather integrated and adapted to form a new theory.[23] The theory reduction options of eliminate, adapt, or integrate are repeated with questions of the body and mind needing to follow the same process. None of these options are in conflict with the MUP, as they all emphasise that the scientific process and its conclusions (the physicalist stance) are correct, irrespective of eliminativist, adaptive, or integrative outcome.

Concerning multiple realizability, Churchland (1986) asserts that the reductive agenda is not in conflict with multiple realizability and is seen as a common feature of science (e.g., temperature is both reductive and has multiple realizability). Furthermore, Churchland (ibid.) acknowledges that mental states and features may be realized within silicon substrates. Multiple realizability can be seen, from this perspective, as a description of a physical outcome that may have multiple avenues to achieve it; a convergence phenomenon with multiple pathways leading to the junction such as the many roads that lead to Rome (section 2.6.3. further explores this).

In relation to the reliability of folk psychology as an explanatory theory, there are multiple features of neuroscientific research that may call it into question. For example, consider confabulation research, where confabulation can be defined as the sincere, or unintentional, reporting of a false memory and is distinguished from malingering where the false reporting is intentional to mislead for secondary gain (Brown, Huntley, and Morgan, 2018). The confabulator can be seen to construct a reason that is sincerely held but inconsistent with observed phenomena. One context of confabulation is in psychological pathology within split brain patients (e.g., Gazzaniga, 1995, 2005), where the verbal hemisphere is separated from the non-verbal hemisphere. The patient is presented with certain stimuli only observable to the left or right side of their visual field. Data that are available to the non-verbal hemisphere is not readily available for verbal report, but if asked what could be done with the object, the person responds appropriately. For example, the non-verbal

23 In modern physics, the ongoing need to integrate disparate theories may be exemplified in string theory (Kaku, 2012) and loop quantum gravity (Rovelli, 2008) as competing theories to integrate general relativity and quantum physics.

hemisphere may be presented with the picture of a toothbrush. When asked what is within the field they are unable to verbally report but when asked what they can do with the object they may produce the brushing motion. If then asked why they are performing the brushing motion, the person may confabulate and state, “I am just stretching my wrist” or some such other fabricated narrative.

Confabulation is also present in non-pathological psychology. Nisbett and Wilson (1977) performed various research projects, which indicated that verbal response (introspective self-report) contrasted with probable causal processes that lead to behaviour. For example, students were asked to remember paired words, such as ocean-moon (here attempting to cue the word “tide”). Later they were asked to respond to apparently unrelated questions to see whether the cued term would appear more readily. Regarding the ocean-moon pair, the researchers asked what the subjects favourite detergent was and the brand “Tide” was preferred (double the frequency with a p value of <0.001). When subjects were asked to explain why they had chosen Tide detergent, none of the subjects referred to the paired associated words but produced reasons such as “my mother uses Tide” or “I like the Tide box” and so on. What these experiments indicate is that folk psychology, as an explanation for behaviour, may be of limited use and external variables (external processes) based on empirical research may offer more reliable explanations.

Furthermore, scientifically valid internal processes also call into question the reliability of folk psychology as an explanation of behaviour. For example, Libet's experiments (e.g., 1985, see also Libet, Freeman, and Sutherland 1999 for further discussion) contrast the neural processes involved in decision making with the conscious experience of decision making. Participants were asked to make a simple decision (blinking, flexing of wrist, and so on) and to time when the conscious mental decision was made (by reporting the time of decision recognised by a nearby clock). At the same time, neural readings (through electroencephalography) were conducted. What was discovered was that identifiable neural patterns (sufficient to predict the decision to be made) were observed that preceded the conscious decision. The implication of this research is that some decisions (at the least in part) are made prior to the consciousness of the decision and by neural activities that are beyond the eyes of introspection.

The above-mentioned research has not settled the reliability or unreliability of folk psychology as an explanatory theory (the research may indicate outliers, refer to only unreliable aspects of folk psychology, or other interpretations given), but these findings do suggest that the reliance on introspection may be a questionable methodology. In essence, this indicates that the scientific image is to be preferred over the manifest image. As eliminativism holds that the mind is the result (however described) of standard physical processes and that most eliminativists agree

that physical processes are multiply realizable (and least in the sense of the concept of multiple realizability defined earlier in this chapter and further developed in section 2.6.3), the MUP is feasible in the latter case if those physical processes are sufficiently replicated.

2.5.2.3 Illusionism

Frankish (2016) presents his version of illusionism as a physicalist representational functionalist stance (see section 2.6.2). Illusionism aligns with preference for the scientific image over the manifest image and with the claim that science (neural mechanisms that relate to anatomy, cognition, and so on) is the preferred method for determining ontological realities. Phenomenal properties (qualia, WIL, and so on), as opposed to phenomenal concepts (representations), are the primary subject of discussion and these phenomenal properties are what is claimed to be an illusion. Frankish (ibid.) accepts that the mental (the manifest image) is anomalous to the physical (the scientific image) and this anomaly may be explained though the acceptance of mental properties as appearing as ontologically real ('things'), yet in reality being an illusion (they are representations of non-real objects).

Illusionism may be associated with Hume's (1777/1975) view of the self (and, by implication, the mind), known as the bundle theory, where the mind (and the self) are not entities but bundles (processes) of experiences. The mind, in this theory, is not a unified entity (a thing) but a coordinating process (how things hang together) that gives the illusion of an entity. Therefore, in both Hume's and Frankish's view, there is no self, only the experience of self, which is an illusion. If the mind is an illusion produced entirely by physical processes, it still falls within the purview of the MUP. In this view all our current and historical notions of mind have been an illusion. Therefore, to mind upload would be to continue the illusion. If the illusion is the product of physical processes and those physical processes can be replicated in another substrate (multiple realizable physicalism), then mind uploading is feasible.

Both eliminativism and illusionism assert that the appearance of the mind is a result of standard physical processes and, therefore, they may be depicted in terms of levels as Ll → Hl because the lower level standard physical processes cause the appearance of higher level (psychological) processes without the higher level having any cause on the lower level (the higher level is an illusion, an epiphenomena, and may be eliminated as a causal explanation theory).

2.5.2.4 Intentional stance

In Dennett's "intentional stance" (1981; 1988b, 1989), there are three stances relevant to the mind placed at various levels of abstraction namely; the physical stance, the design stance, and the intentional stance, each aiming to predict behaviour at the various level of interest. In the physical

stance, the behaviour of the system is understood in terms of precise cause and effect and falls primarily under the domain of physics and chemistry. The design stance looks at the functions of the system and can "ignore details of physical implementation" (ibid, p. 496) and falls under domains such as engineering and biology. The intentional stance is ascribed to the level of minds (psychology) and can be attributed to any system "whatever its innards" (Dennett 1988b, p. 495), where the behaviour of the substrate can be predicted and defined by attributing intentions (e.g., desires, and beliefs, that motivate actions – the DBA model). In this sense, the intentional stance is a predictive descriptive heuristic that occurs when a system (which may equally be described at physical and design stances) is sufficiently complex such that ascribing intentions is pragmatic to predicting behaviour.

The intentional stance abstracts assumes the entity attempting to be understood (predicted) is rational and predicts behaviour based on this stance. The intentional stance may be applied to software and minds and is associated with folk psychology. That minds and software are referred to as both being intentional kinds indicates a computational theory of mind. Elsewhere, Dennett states that "your individual consciousness is rather like the user-illusion on your computer screen" (2017, p. 346-347) and is, therefore, aligned with illusionism as well as computational representational theories of mind. The intentional stance reflects the current thesis' distinction between higher and lower physical levels, as well as the multiple realizable physicalism needed for mind uploading. If the mind can be designed (the design stance) resulting in an intentional stance (the stance that it is commonly associated by persons with a mind), then it can be achieved so long as the system is functionally isomorphic. It is also multiple realizable as it refers to 'any system whose performance ... irrespective of innards' as well as being reductively physical (each system depends in some way on the physical level). Similar to identity theory, the higher level (psychological) is an alternative description of the lower level (e.g. standard physical activity) and, therefore, this view may be depicted as Ll = Hl.

2.5.2.5 Embodied cognition

Shapiro edits a collection of works in *The Routledge handbook of embodied cognition* (2014) that emphasises that cognition (the mind) is probably a whole substrate endeavour (i.e., not limited to the brain or nervous system and potentially the whole body) that occurs in an environment. Both the whole substrate and the environment may be deemed to be constituted by standard physical properties and standard physical processes. Within this section, the emphasis on embodied cognition relates to the whole body (whole substrate) rather than the environment. Embodied cognition, therefore, may be seen as opposed to the neurocentric view but not opposed to the MUP (as

acknowledged by Cappuccio, 2017), and may include internal (here including further whole substrate processes beyond the brain) and external processes (such as behaviours). These whole substrate processes may still be determined as Ll (e.g., stomach processes may relate to feelings of anxiety) as the emphasis is on standard physical biological processes. The difference between the embodied cognitivist and the neurocentrist is that the neurocentricist limits the internal processes to neural activity, whereas the embodied cognitivist expands the internal processes to include other biological activity.

If embodied cognition is correct, then the MUP shifts from a "whole brain" emulation towards a "whole body" emulation (a detailed robotic or carbon copy shell that would need to be developed alongside brain emulated systems). It is an engineering problem as to what elements of the system need to be replicated rather than a metaphysical problem of whether a mind can be a replicated artifact. In this sense, multiple realizable physicalism may be retained, while the extent of the system needing to be emulated may need to be further explored as the project develops. Within this thesis, the neurocentric view is supported as discussed in thought experiments, such as the unlucky soldier (section 2.5.2.1), however embodied cognition (as a denial of the neurocentric view) holds no challenge to the metaphysics of replicating a physical system and the current thesis accepts this as one possibility for the MUP.

2.5.2.6 Evaluations of standard physical Ll internal processes

This section developed further one of the themes developed under external processes in asserting that the mind is a process (with the focus in this section on the internal) rather than an entity. This may be seen as variation on what the substrate does, only this time the emphasis is on what parts of the substrate do internally. The section has focused on internal processes that occur with a standard physical substrate (designated as Ll in this thesis) where the properties of the substrate are standard physical properties, such as already established in the sciences (e.g., energy, mass, neurons firing, and so on). Asserting that the mind is a process may account for why the manifest image of the mind has disparate features to the body. For example, as presented in the corpse problem, the mind is said to cease yet the body remains. If the mind is understood as a process (or set of processes) of the body, then this distinction can be resolved asserting that the mind is linguistically separable (we can speak of the body existing without a mind being present), yet causally inseparable (there can be no process without a causal entity to instantiate it). Representations are further addressed as Hl-processing in section 2.6.2.

To resolve the mind–body problem, the sections on internal processes presented options of either Ll = Hl (identity theory) or that Ll → Hl (illusionism, eliminativism, intentional stance).

For the MUP, both of these indicate that what is required for the MUP is the replication of the Ll (defined here in terms of standard physical properties and processes), and because it is evident in modern science that standard physical properties are realized in multiple locations (multiple realizable), the MUP is, therefore, metaphysically feasible. Other internal processes that have been evaluated are that of embodied cognition, which, if true, would increase the part of the substrate that needs to be replicated; and the extended mind theory, which allows for an integration of internal and external processes (as the external environment is retained in the MUP, a replication of the internal should be a sufficient condition for mind emulation).

Whether a synthetic artifact or biological artifact (the A and B of the ABC options) is needed to replicate these Ll properties and processes awaits further empirical research. What is significant for the MUP, is that Ll properties and processes are that of a standard physical substrate, and that standard physical substrates are multiple realizable (all may in principle be replicated) and, therefore, the MUP would be metaphysically feasible.

The current thesis now suggests a new Turing test that incorporates both external and internal processes (a Turing+ test) where both the behaviours of the organism and the "behaviours" (processes) of neurons could be measured. In the Turing+ test, a correlation can be observed between the external and internal processes (at whatever level of gradation that may be necessary). For example, neural processes could be observable (which parts of the brain react at certain times, in certain sequences, with certain strengths of activation) in correlation to the traditional Turing test's external behavioural processes. If the same processes, both internal and external, can be observed in a second substrate (e.g., a connectionist neural network and a robot interacting with the environment), then it could be said that they have an identical process. If they have an identical process, then they can be considered to be the same mind (based on the assumption that the mind is both an external and internal process).

2.5.3 Questions for the standard physical substrate view

Presented here are two common thought experiments used as objections to physicalism that affirms standard physical properties (standard physical substrates), with an emphasis on the limitations of the current scientific method and processes that are applied to an understanding of the mind.

2.5.3.1 What is it like? (Nagel)

Nagel (1974) presented a thought experiment where he asked what is it like to be a bat? For Nagel, the physicalist agenda (here referring to those that would assert that the standard physical substrate is sufficient) within philosophy of mind is limited and requires further integration of the subjective

experiences such as consciousness (see section 2.7.1 regarding Chalmer's 'hard problem') and the need to broaden our scientific understanding of the mind to include subjectivity. As such, it is largely a critique of the reductive theorists (identity theorists, illusionists, and so on) and an argument to incorporate phenomenological features into the physical domain (further physicalism). The emphasis is on the manifest image, the what-it-is-likeness (WIL). Nagel (ibid.) begins with the assumption that the subjective first person experience cannot be understood through the explanation of functional analysis (the physical causes) because then these could be ascribed to automata that perform the same functions, yet do not have the same experience (similar to philosophical zombies, further developed in section 2.6.4). He acknowledges (in a footnote) that perhaps a sufficiently complex automata could have experiences, but that this cannot be asserted based on analysing experiences. The uniqueness and validity of first person subjective experience is what Nagel is attempting to emphasise and the general argument can be posited as follows:

(1) Bats have first person subjective experiences.
(2) Physical descriptions (here suppose a perfect scientific knowledge of the bat's physical functions) would not provide an understanding of this experience.
(3) Therefore, physicalism fails to address the first person subjective experience.

This thought experiment is then applied to persons and the mind. For Nagel (ibid.), if a perfect neuroscience of the mind were to exist, it would still leave out the first person subjective experience (the WIL) and, therefore, the physical scientific process (although valuable) is insufficient and requires the incorporation of WIL (which is something over and above the physical sciences currently understood) to the philosophy of mind.

For Hofstadter (1982a), Nagel's what it is like argument is seductive but Hofstader is not convinced that it "makes any sense to project ourselves onto the object in question" (ibid., p. 406) and whether this projection really refers to anything at all. Hofstadter (ibid.) gives a multitude of WIL scenarios "What is it like to believe the earth is flat? … What is it like to be a running AI program? ...What is it like to be the Rock of Gibraltar" (ibid., p. 404-405). Nagel (1974) is aware that not all entities may be conscious and chooses a bat because it is similar enough in constitution for most to attribute a conscious process, yet dissimilar enough for the WIL to be beyond our projective understanding. Hofstadter's (1982a) point is not to argue that the Rock of Gibraltar is conscious but to argue that WIL is subjective in the sense that only the system instantiating the experience can experience it (this is the definition of a subjective experience). WIL is, therefore, the

process that occurs when a particular physical system instantiates a functional process and is, therefore, an indexical process. The question of WIL is a projection based on assumed similarities between the indexical system and the system of projection (e.g., I can imagine what it is like to be famous because the state of fame is similar to previous experiences where this status was achieved). The less similar the system, the less accurate the WIL, as will be the case with bats.

Queries emerge in relation to the MUP as to whether an artifact can be similar enough for the experience to be sufficiently synonymous to be an extension of the self. First, as our own experiences are dissimilar throughout time (the experience of the self alters with age, location, and many other variables), there only needs to be a similar enough instantiation rather than a perfect one (see Goertzel, 2012, for one such possibility based on fuzzy state systems and probabilistically weighted transitions). This relates to the question of continuity, or persistence, of persons and is the subject of the Chapter 3. Furthermore, the question is whether standard physical processes can account for the first person subjective experience. This is the fundamental mind–body problem from the standard physicalist perspective. As discussed within the current chapter, there are various avenues of response. If this criticism of standard physicalism is upheld, an emergentist or panpsychic solution may also be sought that may maintain the metaphysical feasibility of the MUP (see section 2.7), or alternatively, this criticism may be viewed as insufficient to cause difficulty for those who assert a standard physicalist perspective. The current thesis is inclined to maintain the view that subjective experience is the indexical physical process of a functional system. To project WIL, therefore, depends on the similarity of the indexing system to that of the projected system. In this view, the more similar the system, the greater the correlation of WIL and, furthermore, if the system were functionally isomorphic then WIL will be the same.

2.5.3.2 The knowledge argument (Jackson)

Jackson's (1982, 1986) knowledge argument can be introduced through the thought experiment of Mary the neuroscientist. Mary has lived her life in a black and white room with only access to black and white images of any sort. In this room she has become the world's leading neuroscientist of vision obtaining a *complete* knowledge of the physical aspects of colour (e.g., what wave lengths each of the colours occupy, how the retinal cones respond when activated, the precise patterns of neural correlates when seeing particular colours, and so on). In this she is said to have the complete physical information of colour and its perception. Then, one day she is released from the room and *sees/experiences* colour for the first time. Has Mary learned any new information? If yes (the assumption of Jackson here), then new information is obtained that is not available in physical information (which was complete prior to exiting the room). If new information is achievable through

first person experience that is not available through physical information, then it can be said that physicalism (standard physical properties and processes) is false (or an incomplete metaphysical stance). The argument can be simplified in the following form:

(1) Mary has complete physical information prior to release.
(2) On release, she learns something new (acquires new information).
(3) Therefore, physical information is not complete information (physicalism is false).

As Prinz (2012, Chap10) notes, one of the difficulties with the knowledge argument is that it attempts to make a metaphysical claim from an epistemological argument. In this sense, the argument may, at best, express limitations on the physical knowledge that we can obtain rather than denying that everything is physical. If it is assumed that Mary can immediately identify red when Mary steps out of the room (a contentious notion because many experiences involve learning within a social context, such as a mother repeatedly identifying colours), what happens? The physicalist can assume that new neurons fire and that new parts of the physical system are activated. Mary may be aware that red activates neural pattern y prior to exiting the room, but this discursive knowledge will involve the activation of neural pattern x and is not the same as activating neural pattern y. Therefore, the activation of neural pattern y (when she sees red) will be a new physical process and physicalism may be defended in this regard. As Mary has prior information that activation of neural pattern y leads to the experience of seeing red, what varies is not the information (that neural pattern y is activated) but the instantiation of neural pattern y. This, therefore, queries one of the assumptions, namely; that Mary *learns* something new. What would happen if while in the black and white room Mary were to stimulate the neural pattern y? Would she not have the qualia of red? If this is affirmed (that she would see red), then the distinction between physical knowledge (as defined in the knowledge argument) is one of instantiation of the appropriate system (with both the discursive knowledge of the system and what is required for instantiation to be within the purview of science).

The above response is similar to previous critiques where what is *learned* (the qualia experience) can be a new ability (Lewis, 1988), a know-how not knowledge-that which is a non-propositional acquaintance knowledge, an old-fact (physical information) as a new representation (the qualia), and so on (see Williamson & Stanley, 2001; Nida-Rümelin & Conaill, 2019). The main idea here is that the type of knowledge Mary learns on exiting the room is not new theoretical or discursive knowledge but a new presentation of it. If it is assumed that premise (1) is feasible (a contentious notion because how completeness is defined and what completeness would entail are

unclear), then the distinction between (1) and (2) can be seen in variation in knowledge of a functional system and the instantiation of that system. A schematic of an electric circuit (here imagine all the functions of every lower level, such as conductivity, included) will have all the necessary and sufficient knowledge of what is needed but will not light up the room. Only once that knowledge is applied to a physical instantiated system (i.e., someone builds the circuit as described), will the light be produced. If this is applied to the knowledge argument, the qualia can be seen to be the result of physical instantiations (they occur only in an active system), yet nevertheless be functionally defined (whether Hl or Ll). If this is the case, the qualia can either be causal or non-causal. If they are non-causal the qualia can be described in the formulation Ll → Hl (e.g., as in illusionism) and if they are causal they can be formulated as Ll ←→ Hl (e.g., as in property dualism).

For further discussion on the knowledge argument, see *There's something about Mary: Essays on phenomenal consciousness* and Frank Jackson's knowledge argument (Ludlow, Nagasawa and Stoljar, 2004).

2.5.3.3 Evaluations of questions for standard physical substrates

The above criticisms of physicalism have emphasised the type of physicalism that has a standard physical substrate. The argument for multiple realizable physicalism (section 2.2) argued that should a mind be physical, it would likely be multiple realizable even if the substrate were some further physical substrate (see section 2.7.) and, therefore, that should these criticisms be upheld, the MUP would still be feasible (albeit instantiated by a further physical substrate). As demonstrated above, there are multiple avenues for the proponent of the standard physical substrate to counter these criticisms. For example, qualia/WIL may be rejected as being real (Dennett 1998a) and being an illusion (Frankish, 2016). Qualia/WIL may be accepted as real but not as being an essential condition. For example, Helen Keller (Keller, 1905) lost her sight and hearing at an early age, yet this seems to have had little impact on her consciousness or sense of self, as is the experience of many persons who loose faculties. The current thesis views qualia/WIL as simply what happens when an indexical system represents its own internal processes and is developed further in the novel contribution of the concept of the efferent-self (see section 4.3). In this view, the knowledge of the system (scientifically defined) is not the same as the process of activating the system (e.g., knowledge of light does not light up a room) and WIL is a projection of the indexical system to systems similar to itself.

2.6 Functionalism

Having presented theories presenting the mind in terms of internal and external processes from a primarily standard physical substrate perspective, the current thesis now turns to functionalism which, as will be demonstrated, allows for alignment with multiple physicalist philosophical perspectives (from those who affirm a standard physical substrate to those who affirm a further physical substrate). Unlike the illusionist and eliminativist, the functionalist accepts the manifest image of folk psychology (or, at the very least, aspects of it, such as in the DBA model) and attempts to unify this with the scientific image (largely with neuroscience) and so hopes to retain and integrate both theories. The functionalist strategy aims to effect this integration by referring to a third theory (offering a neutral description[24]), which describes the causal roles that can be applied to the mental in both folk psychology and neuroscience. Within functionalism, mental state types are asserted to be the same as physical state types if both can be described in functional terms. To do this, a Ramsey-sentence (e.g., Schwarz, 2015; Levin, 2018) may be invoked as a neutral description for both folk psychology and physical sciences using functional analysis. Ramsey-sentences, attributed to the late Frank Ramsey (1929), were so titled by Hempel (1958). The main concept is that statements in different theories (e.g., the desire to eat ice-cream in a psychological theory and the correlated neural activity in a neuroscientific theory) can be united if generalisations are conjoined replacing descriptions of mental types with variables and then to quantify existentially over those variables. For example, Levin (2018, section 4.1., para. 3) presents pain as "∃x∃y∃z∃w (x tends to be caused by bodily injury & x tends to produce states y, z, and w & x tends to produce wincing or moaning)." It is noted that in the above example there are no mental state terms such as pain yet the quantifiers (e.g., tends to cause bodily injury, wincing, and so on) specify any physical observable causations that are commonly designated to mental states. The use of functional terms may, thus, be seen to integrate mind/body terms.

In neuroscience, textbooks statements such as the following are common:

> "What we commonly call the mind is a set of operations [functions] carried out by the brain" (Kandel *et al.*, 2013, p.5), which led Campbell (2018), in a lecture,

24 Functionalism can, therefore, be seen to be similar to neutral monism, double-aspect theories, logical behaviourism, and any other such theory that attempts to solve the mind–body problem through defining a third (alternative) way to discuss what appears disparate.

to state in relating philosophy to neuroscience that functionalism is “the only game in town”.

Therefore, within the physical sciences of the mind functionalism carries the empirical trajectory.

Functionalism, broadly speaking, is the view that the mind is identifiable (with varying degrees of strictness) with the causal functions of the system (however defined) that are associated with the mind. What matters for the mind is not what the mind is constituted by (e.g., biology or silicon) but what functions are performed in relation to both external (sensory stimuli and behaviour) and internal (other mental states) processes. If it is these functions that are necessary and sufficient conditions for the mind, then the nature of the substrate that instantiates the functions are not essential and the mind is multiple realizable. Consider two slinky toys, joined together, one made of metal and another made of plastic. Now move the one end of the slinky up and down creating a wave function. The wave function moves across the one substrate (e.g., plastic) to the other substrate (e.g., metal). The function requires a physical instantiation (either the plastic or metal substrate) yet is able to move across substrates. It is, therefore, substrate-independent but still substrate-supervenient (it does not exist without a physical instantiation). Physical functionalism, in relation to the philosophy of mind, states that the mind is akin to the wave function in the above example. The current thesis presents three functional premise which are:

(1) The functions are the mind (metaphysical identity claim).
(2) Knowledge of these functions is necessary and sufficient to understand the mind (epistemological claim).
(3) Not all substances/properties can perform the same functions (nomological/empirical caveat).

The first premise (1) does not commit functionalism to a physicalist stance (e.g., the functions may be the functions of the soul or some other non-physical causal process), but it is primarily adopted by physical monists (e.g., Smart, 1959, 2017; Lewis, 1986) and that is the stance adopted by the current thesis. Functionalism emphasises causal roles and defines these as the necessary and sufficient conditions for understanding the mind and because identity theory also emphasises causal roles these theories are closely related (Smart, 2017). The functions can be seen as processes and, therefore, de-emphasise the metaphysical nature of the entities that perform this process. The function may, therefore, be applied to various substrates and have multiple instantiations. Advocates of

functionalism provide many artifact examples, such as mouse traps, electric circuits, and so on, to illustrate that the constitutive physical matter (what physical matter makes up the artifact) is not constitutive of the functions, as the same design (the mouse trap, the electric circuit) may be instantiated with variation in physical elements. In short, whatever does the same thing is the same thing.[25]

Premise (1) may be queried as to whether functions are sufficient and whether the mind may include something further (e.g., WIL, qualia) to these functions. This difficulty leads to the second premise (2) that knowledge of the functions will lead to a sufficient understanding of the mind. The physicalist who denies this has the option to assert that further aspects of the mind are either something inherent in physicalism but not easily accessible through functional scientific analysis (see section 2.7.1), or that these aspects emerge as further properties (see section 2.7.4). Objections to functionalism and the assertion of some further phenomena beyond the function as needed for a mind is the subject of section 2.6.4.

The third premise (3) includes a nomological caveat to the functionalist discussion and asks can anything perform the same function? If any substrate can perform the function it is the same thing, but is this practical from an engineering perspective? Could the wave function of the slinky be performed with a brick or can we perform flight with a liquid wing? These would make the engineer's (artifact creator's) task, at the least, impractical and, at the most, impossible. Therefore, the metaphysics of functionalism (the claim that anything that performs the same function) has a nomological boundary, in that different physical substrates have different properties that may impact on the function (i.e., not all things can practically be used to perform the same function).

Premise (3) may also relate to the issue of levels and emergence relating to functionalism (Lycan, 1998), in that each level of a system implies a lower level of functional analysis. For example, in an electric circuit, a very coarse schematic may be drawn (a functional description), but within this schematic, further functions of the components are assumed (e.g., the function of conductivity). The function may be viewed as a type (the electric circuit design) with the physical instantiations (each physical electric circuit) as the tokens and so reflects the type/token distinction. Within each of these tokens, a further functional type/token distinction can be made (e.g., the functions of conductivity, atomic bonding, quantum fluctuations, and so on). Therefore, in determining

25 The use of the term 'thing' here illustrates how it can be used as a process ('does the same thing'), or an entity ('is the same thing').

functions it may be needed to understand the lower levels, or at the least have sufficient heuristic knowledge of lower level properties, to allow for the higher level functional analysis.

Levels of functions here are again referred to as higher level functions (Hl) and lower level functions (Ll), with Ll functions relating to neuroscience (e.g., neural firing, chemical interactions, and so on) and are, therefore, based on standard physical properties and processes, whereas Hl functions relate to psychology (e.g., cognitions, emotions, and so on). The Hl functions may refer to standard physical substrates performing more complex functions than the Ll processes ('higher' here meaning more complex functions that are reducible to Ll functions) or to further physical substrates ('higher' relating to emergent properties and their processes). The nomological boundary (see section 1.5.6) indicates the necessary and sufficient conditions for a functional system to produce the desired outcome (the function/s that are aimed at being replicated). The mind–body problem primarily focuses on the nomological levels of the biological and the psychological (the primary distinction within the problem). Therefore, further lower nomological levels (such as quantum mechanics[26]) are not emphasised here.

If external processes are included, a formulation such as the following may be presented, Env←→ Ll ←→ Hl. Lower functions here are the physiological processes (e.g., neurons firing) that interact causally with higher level functions (e.g., psychology) and are reflections of what the substrate does internally. In its turn, behaviours are the functions of the substrate whole as it interacts with the environment. Functionalism may, therefore, include both internal (e.g. Putnam, 1965, Lewis, 1972) and external processes (e.g., Harman 1990) constituting the mind, if it is deemed that they both have causal roles:

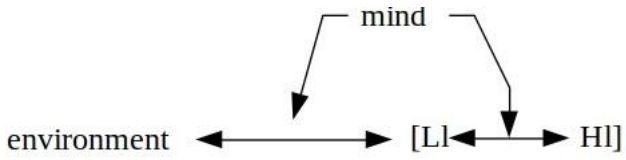

Figure 2-3: Functionalism as external and internal processes

26 Penrose (1989) has championed some movement to include even lower levels, such as quantum mechanics, however, this thesis is unaware of any quantum activities that are not sufficiently explained at the biochemical level and, therefore, the nomological level of quantum (although present) need not be included. However, even if quantum were relevant, this would greatly increase the complexity and difficulty in sufficient reproduction (an engineering problem) but offer no metaphysical objection.

If the external functions only are included, this is akin to the external processes emphasis of section 2.5.1 with the functions as causal roles (here represented by arrows).

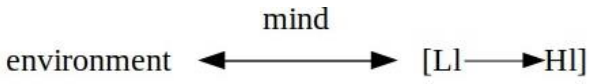

Figure 2-4: Functionalism as external processes (e.g. behaviourism)

Since, in this instance, the mind is identified with behaviour (here external functions), whatever entity produces the same external functions can be said, in this view, to have the same mind.

However, functionalism traditionally (e.g., Putnam, 1965; Lewis, 1972) focuses primarily on the internal functions (Ll and Hl) and, therefore, although functions can in principle be expanded to include functions within the environment (external processes), the current thesis focuses on higher and lower levels (in terms of internal processes) as presented in Figure 2-5.

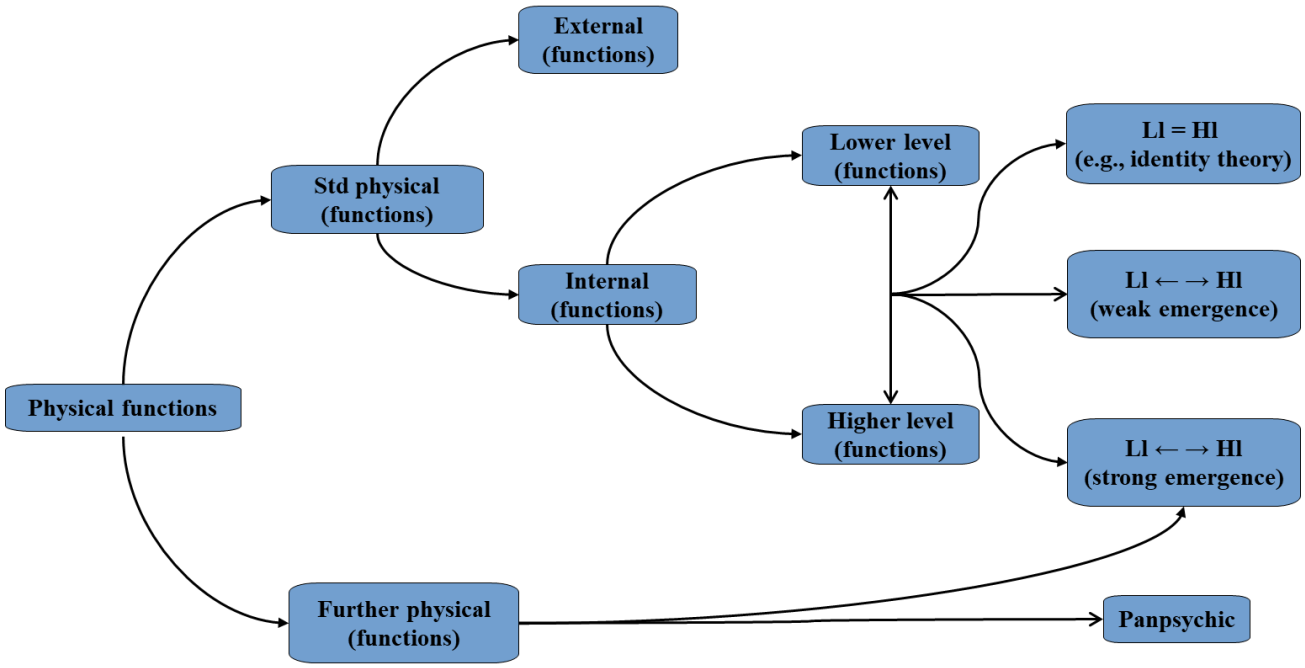

Figure 2-5: Functional physicalism

First, as stated earlier, the focus here is on functionalism that holds to physical monism (thus, non-physical substance theories such as substance dualism are not present within Figure 2-5). Second, the diagram only deals with processes that are causal (the definition of functions). Third, the emphasis is on internal functions that are further categorised into lower or higher level functions (external functions briefly discussed above). Fourth, the interaction between the Hl and Ll for functionalism lead to three options, namely a) the Hl are the same as the Ll and, therefore, akin to some form of identity theory b) the Ll interactions lead to Hl properties, that may be weak emergence (reductive physicalism) or c) the Ll interactions lead to Hl properties, that may be strong emergence (non-reductive physicalism). For options b) and c), the higher levels are causal in their own right and, therefore, are part of the functional system (which is defined as the mind). Within the strong emergence, a further physical property has emerged. As this further physical property is non-reductive, it is associated with a further physical substrate (mind will be instantiated in mental properties that emerge at that higher level). Panpsychism is presented in Figure 2-5, although it is not emphasised within the current functional section but will be presented in relation to the work of Chalmers (section 2.7.1). The panpsychist is distinguished from the emergent views (whether weak or strong), because the panpsychist's further property is not based in lower level standard physical properties. The current section focuses on emergence (weak and strong) that has standard physical basal properties.

For the MUP, what is aimed at is achieving the Hl that represent the psychological mind (whether identified with functional causes or not). If Hl is deemed to be causally irreducible (but still physical), then what is necessary is that replication occurs at that higher level (similar to if the solidity, as a higher level property, can be used to make tables from any substrate, then it is that Hl property that is defined as necessary and sufficient in relation to the substrate properties). The irreducible Hl, therefore, designates a sufficient level of replication (nomological boundary). If the Hl is reducible, then this allows for the possibility that Ll would also need to be replicated (as Ll also forms part of the causal functions), which increases the engineering complexity of the MUP but not the metaphysical difficulty.

2.6.1 Forms of functionalism

Within functionalism, there are multiple forms. Block (1996b) identifies three types (functional analysis, computation-representation, and metaphysical functionalism), whereas Levin (2018) identifies four (machine state functionalism, psycho-functionalism, analytic functionalism, the role versus realizer functionalist distinction). The current thesis categorises functionalism under two

broad categories that emphasise higher level functions (Hl) and lower level functions (Ll), as this reflects some of the categories within the literature.

Those who emphasise Hl, emphasise that the nomological (scientific) domain of psychology is somehow causal. For Block (1996a, 1996b), computer-representation and metaphysical functionalism, and for Levin (2018) machine state functionalism, psycho-functionalism, and analytic functionalism, are theories that argue that higher level functions are to be identified with the mind and that they play a causal role (also called role-functionalism).

Block's (1996b) category of metaphysical functionalism and Levin's (2018) category of analytic functionalism are largely synonymous. They assert that these categories reflect a form of functionalism (the role of causal functions) that asserts a metaphysical *a priori* necessary truth applicable to all possible worlds (e.g., Lewis 1966; Armstrong, 1968) [27], while in contrast, other identity functionalists, such as Place (1956) and Smart (1959), have a contingent stance. In this view, any being that has the same functional isomorph is, therefore, functionally equivalent and instantiates the same mental state. A mind *is* a process that *performs* mind (mental causal) functions. Whether proponents hold to the mind being *a priori* or contingently functional, both groups assert that the mind is to be identified with physical functions. As the psychology (Hl) is synonymous with the functions of the Ll, in this view, it would be necessary and sufficient to replicate the Ll (neuroscientific functions) to replicate the mind. Furthermore, because Hl is identified with the Ll functions, both are causal as they both are the same physical processes.

But what is the nature of these higher level functions? Commonly (further developed in section 2.6.2), the higher levels are defined in terms of information processing such as computation (Block's 1996a, 1996b – computer-representation; Levin's -2018 – machine state functionalism categories) and representation (Block's 1996b – computer-representation; Levin 2018 – psycho-functionalism). These information processing processes are commonly deemed to be causal functions in their own right and are, therefore, emergent properties (either weak or strong emergence).

Levin (2018) notes that a distinction that cuts across the functionalist literature is between role-functionalism and realizer-functionalism. According to Levin (2018), realizer-functionalists

---

27 It is worth noting that these authors are also proponents of identity theory and type physicalism, indicating that functionalism is consistent with the reductive identity theories (as developed by Jackson, et al., 1982, Smart, 2017; Akand, 2018) For example, Armstrong (1968) defends his reductive type identity view by asserting that mechanisms (here meaning computers of a sort) will soon be creative and that once this is done, "it will be very arbitrary still to maintain that there are expressions of mind that are beyond the powers of mechanisms to produce" (ibid., p. 357). The mind is, therefore, something that may occur across substrate kinds.

assert that functional theory is rather a description of lower level properties. In this sense, the higher order functions are not causal properties but rather descriptions (abstract concepts) that themselves serve no causal function but, nevertheless, refer to the lower order physical functions that are themselves causal. In this view, the mental event of pain is not a metaphysical entity in itself but rather a non-causal description of the Ll functions. In this sense, the functions that matter are the neural processes, whereas as the psychological descriptions (e.g., pains) are non-causal representations of these Ll processes. In contrast, the role-realizer asserts that the higher functions (such as pain) are metaphysical states in their own rights.

For Block (1996a), this distinction between the lower level standard physical activities and higher level representations (Hl) are defined using terms of the functional state identity claim and the functional specific claim. In functional state identity, the claim is that the higher order mental function *is* the property of being in a certain state. For example, pain is not identified with the physical state but rather the function that is performed by lower order physical states. The function is, therefore, a property in its own right, a thing (entity). In contrast, the functional specific claim moves away from identifying the abstract function and closer to the physical specifics of a pain event. Pain, in this view, is only an abstract description of multiple physical tokens, but it is the physical tokens that are identified with pain (the physical processes that are causal functions). Block (ibid.) uses an example of a car carburettor where the functional state theorists would assert that a "carburettor = being what mixes gas and air" (ibid., p. 38), whereas the functional specifier would say that "the carburettor is a type of physical object". The functional state identity asserts that the causal function is the functional description, whereas the functional specifier would state that the causal function is a description of specific physical instances.

Papineau (2000), in describing functionalism, mentions first-order functionalism (the physical state that actually plays the causal role) or second-order functionalism (where it is "the state-of-having-some-state-which-plays-a-certain-role"). First order functionalism is, therefore, akin to realizer functionalism and functional specific claims, in that they all emphasise lower level physical functions. In these views, the Hl functions are descriptions of Ll causal roles and may be seen as the score keepers of the causal game, identifying what happens but playing no role themselves. Alternatively, the role-realizers, functional state identity, and second order functionalism see the Hl functions as part of the causal game (perhaps an umpire, the boundaries of the field, the rules of the game, and so on).

Table 2-2 categorises the variations of emphasis discussed within the functional literature under the two primary categories of higher level and those who emphasise the lower level.

Higher level	Lower level
Second order functionalism	First order functionalism
Role functionalism	Realizer functionalism
Functional state identity theory	Functional specification claim
Analytic functionalism	
Metaphysical functionalism	
Machine state functionalism	
Representationalism	
Higher order theories	
Psycho-functionalism	

Table 2-2: Functional emphasis of Hl and Ll

Much of this discussion refers to variation in levels of description and levels of causation as they relate to metaphysical claims. What is important to note in relation to the MUP, is that both Hl and Ll (as presented here) are multiple realizable physicalist stances. Both assert that functions require physical instantiations to produce or realize the functions. Both assert that it is the functions (irrespective of preferred nomological level) that are identified with the mind. And both allow for the possibility of replicating the relevant level of functions. The distinctions hinge on the necessary and sufficient nomological boundaries identified with the mind.

2.6.2 Higher level functions as information processing (Hl - processing)

Hl functions may be seen as properties or as information processing[28], with the latter being the focus of the current section. Hl information processing is typically designated as representational and computational theories of mind, which, according to Pitt (2018), are largely synonymous. Is information/representation/computation (Hl-processing) something over and above the standard physical properties and processes (i.e., do these Hl functions result in an emergent property) and what are the implications for the MUP? Consider the analogy of a software program as a type (e.g., a particular word processor). In describing the program, it appears to have different properties as that of the hardware. For example, the software (i.e., the information), has no mass, occupies no

28 Information theory (as it relates to computation) relates to the work of Shannon (1948) where the original context was in relation to efficiency and communication between artifacts. The efficient information system predicts through probabilistic decisions (a form of Bayesian inference) through the use of bits (in modern computation relating to the base binary code of a 0 or a 1), which is the smallest amount of information that needs to be communicated (communication being the function). To note quantum computing increases the complexity (using a qubit as quantum information) but the general concept of information remains. Information theory is now used within biology (e.g., Dawkins, 2003, Chap 2.3) as well as neuroscience (e.g., Staelin and Satelin, 2011; Churchland and Sejnowski, 2016), which furthermore has a specific discipline of computational neuroscience (e.g., Trappenberg, 2009).

time and space (if it is not instantiated in a hardware system). It is an abstract entity. Furthermore, the software engineer does not need to be aware of the composition of the hardware as they design the functional causal role of the software. The software engineer, therefore, constructs, maintain, and repairs software at a relatively autonomous level (e.g., they need have no knowledge of the chemical compositions of the hardware). This Hl autonomous level (nomological boundary of information processing) does not necessitate that there is a strong emergent property of software (similar to the carpenters woodwork analogy the autonomy may be pragmatic yet reducible).

Assume that the mind is an information processing system (the assumption of Hl-processing). It is acknowledged that we do not have a complete understanding of how this processing is done, however, we are aware of other processing systems (e.g. computers). As both are deemed to be processing systems (with variation of complexity), what holds true for the one should hold true for the other. It may, therefore, be inferred that since other information processing systems (e.g., computer software) may be at an Hl autonomous level yet reducible, then so may the mind. Alternatively, the information processing may be asserted to leading to non-reductive properties and a strong emergence may be asserted (although this is not the view of the current thesis). If information processing were such a strong emergent property and the software engineer were indeed working with strong emergent properties of higher level information, what would the implication be for the MUP? As information processing is a physical process, then the artifact creator (e.g., the software engineer) need only be concerned with this nomological level and as demonstrated in software engineering, these are multiple realizable. Therefore, the Hl autonomy as maintained by the nomological boundary (here Hl-processing) may be either reductive or non-reductive, yet multiple realizable and within the purview of science.

In support of standard physical processes (denial of strong emergence), consider that standard physical processes may occur at multiple levels of abstraction and, therefore, there may be variation of what is termed ‘higher’ levels. Consider, an average bee who has the function of flight. Explored at the level of the singular bee, various physical functions can be described and an explanation of flight produced. Now consider a swarm of bees (of which our normal bee is a part); this produces a different level of function and a different description as bees fly in a cohesive pattern. Is this level of functioning physical? It would appear that it is as there is no need (due to Occam’s razor as well as empirical observations) to postulate a ‘swarm property’, which is distinct from standard physical properties and processes, yet nevertheless the description of swarm functions is at a different level (of complexity) than the level of flight of the individual bee (yet both are forms of flight). In a similar way, higher level functions of the mind can be understood with

variation 'higher' level physical processes. This analogy has many variations in the literature; Minsky (1988) refers to the society of mind, Bateson (1979) the ecology of mind, Hofstadter (1979) to an ant colony, Wiley (2014) to flocks, swarms, and so on. There have also been attempts to mathematically formulate the functions as a correlation between the higher-level functions of the mind and swarm functions (e.g., Trianni, Tuci, Passino and Marshall, 2011). These perspectives are briefly mentioned here to illustrate how a higher level of functions can be construed as variation in physical functions and, therefore, to be compatible with a physicalist stance that begins with a standard physical substrate.

As discussed earlier, machine functionalism, psycho-functionalism, higher order theories, and so on, can be classified as theories depicting mental representations as causal processes at higher functional levels (Hl). Some (e.g., Fodor, 1975 Jackendoff, 2002) emphasise language (discursive representation), while another approach (e.g., Kosslyn, 1980, 2005) favours images (pictorial representations). The exact type of representation (e.g., discursive or pictorial representations) is not essential to the current discussion as all may be described as higher level functions. Levin (2018) mentions psycho-functionalism (e.g. Fodor, 1975), where the functionalist emphasises the empirical work and theories of cognitive psychologists. As a psychological theory can be developed through the functional interactions of psyche that need not mention neuroscience or lower order physical processes, it is deemed an autonomous domain to some extent.[29]. Because higher level functions may be sufficiently theorised without reference to the lower level, according to these views, they are deemed autonomous (whether this autonomy is pragmatic and the level is ultimately reductive or whether the level is non-reductive) and, therefore, this higher nomological levels (e.g., psychology, sociology) is sufficient to understand the mind. Should these nomological levels (which are currently not clearly defined but may be in the future) be replicated sufficiently, so would the mind.

Putnam (1960, 1965) related the mental states of the mind to functional states that can be instantiated in a Turing compatible computer and so his view has been termed machine state functionalism. The computer has a functional organisation where tables that refer to inputs and outputs can be abstracted. If the machine is in S_1 (particular state) and receives I_1 (particular input), it will enact a sequence that results in O_1 (particular output). The mental states are, therefore, in this view,

29 For example, Freud's (1955) Id, Ego, Superego has served as a basis for describing the mind and behaviour without reference to the biological. Yet for Freud, this is not a matter of denial of biology but describing biology at a different level for example "... that is subject to further development through an increase in the number of its neurones and through an accumulation of quantity" (Freud, 1954, p. 354-365).

computational functions, with Putnam (1980) preferring "probabilistic automata" to deterministic ones. The description of states, inputs, and outputs can be seen as a variation of psychological models such as the DBA model, as well as Ramsey-sentences as each expresses an output (e.g., action in the DBA model) as a result of input resulting in an internal processing of sorts that leads to an output. Therefore, beliefs and desires (and other mental states such as pain) *are* mental functions. An aim of these early works (Putnam 1960, 1965) was to offer a more psychologically validating theory than behaviourism, while adhering to science through the integrating of computer theory. If we consider the functional levels, Putnam places the relevant discussion of mind as one step beyond the specific physical structure (it is what the physical entities do that are defined as the functions) and, therefore, a function can be considered an abstract type with physical tokens.

Furthermore, the appropriate level of discussing functions, in this view, is not at the level of the direct physical properties (i.e., the specific neural structure) but at the level of information processing among physical entities. Again, this is an abstraction one step beyond the physical entities to emphasise physical processes (both the machine and the person are physical entities that perform the functions). For machine state functionalism, the type identity theorist fails because their theory is not abstract enough to encompass multiple realizability (see section 2.6.3 below). This form of functionalism, if accepted, would indicate that mind uploading would be feasible and, if the original machine functionalism were accepted, it would indicate that an artificial (machine) substrate would be sufficient (the A of the ABC options).

Higher Order[30] Thought (HOT) theories refer to the mind (primarily here in the context of consciousness), where higher order states are used to distinguish the conscious from the non-conscious. To achieve this proponents (e.g., Rosenthal and Weisberg, 2008) use the transitivity principle where a mental state can be described as conscious when it is aware of that state. At issue here are, therefore, meta-representation processes (they represent representations - see also Carruthers, 2016; for further exploration). In terms of Hl, processing information typically refers to an object other than the information conveyed (information of a cat is not a cat), however, if it is accepted that parts of the information processing refers to information, then a self-referencing system is established (information about information, or representations of representations) and leads to, what Hofstadter (1979) calls, a strange loop. This view allows for the possibility of the self/mind

30 Higher order is distinguished from higher level as higher order assumes a second order property, whereas higher level may refer to higher order properties or higher level activities that need not hold property status. For example, the flight of the swarm of bees is a different function than the flight of the individual bee, yet it need not be a new property (such as the property of solidity that emerges as the result of atomic and molecular interactions).

as an indexical self-referencing system where terms such as ‘I’, ‘you’, and ‘me’ (terms used to designate personal identity and the particular mind) may be meta-representations of other representations (however defined).

The causal efficacy of mental representations (the higher level of the system) is an area of debate which Pitt (2018) relates to intentionality. He distinguishes between intentional realists and intentional eliminativists. Intentional realists (e.g., Dretske, 1988; see McLaughlin, 1991 for criticisms and responses of Dretske's views; Fodor, 1975, 1987) argue for the ontological reality of mental representations and, therefore, align with folk-psychology, where Ll ←→ Hl. Alternatively, intentional eliminativists (e.g., Churchland, 1981; Dennett, 1988a, Ramsey, 2007; Frankish, 2016) argue against the ontological status of representations, where Ll → Hl holds. For intentional realists, the manifest image of our psychological life having causal efficacy (e.g., the DBA model) is a strong intuition that is difficult to deny and implies that any theory of the mind should accept the causal efficacy of the mind. The intentional eliminativists may concur that mental representations have a quasi-existence (in the same manner that a rainbow ‘exists’), but that it has no causal role within the system. Dennett (2017) elaborates this point by using the example of a computer screen as a representational system. When we move the mouse cursor on the screen and interact with icons, it has the appearance of causation but the real ‘heavy lifting’ is done by the hardware (the screen is simply a display). This relates to Ll (e.g., realizer-functionalism, first order, and functional specification), where the higher representation functions are descriptions of the physical processes but in themselves play no role in the causal game.

It is also noted here that the representational properties are not the same as the properties of the object of representation (Harman, 1990). Consider a photograph (the representation) of a cat (the object of representation). The photograph does not have whiskers, fur, four legs, purr, and so on, yet these are properties that are intrinsic to the ontology of cat-ness. The representation is flat, static, and yet most persons would easily identify the correlation between the representation and the object of representation, namely; that the photograph represents the cat (and even a particular cat). Is the representation ontologically real? Here the consideration is not that a photograph (the piece of paper that instantiates the representation) is real, as this representation could be held within other mediums (e.g., a computer screen, a projection on a wall), but whether the information conveyed in representing the cat is real. If it is acknowledged as real, it is nevertheless not the same as the object of representation (it has different properties). The ontological status of representation is a complex philosophical question and the agenda within the current thesis is not to resolve these

but to note that from a physical monist perspective representations are the result of physical processes, that they may be expressed within multiple levels of abstraction, and that these properties are not readily definable by standard physical properties such as occupying mass, energy, and so on. Therefore, the mind–body problem, as established through the apparent disparity between the mind features and body features, may well be a disparity between representations and the physical systems that instantiate them.

As Hl-processing asserts that a representational system is the necessary and sufficient level of analysis for a mind, then what is needed is that the representational process (the nomological boundary of psychology if psychology is defined as representational processing) is to be replicated. However, functionalism also may include an Ll emphasis (or a hybrid which includes both Hl and Ll functions). Whatever the nomological boundary, if it is asserted that these functions are physical (requiring physical instantiations) and multiple realizable (that a particular set of functions may be replicable), then the MUP would be feasible.

2.6.3 Multiple realizability in functionalism

Multiple realizability is an inherent factor of the functionalist agenda, in that if it is the functions that matter, then these functions can, in principle, be produced by other substrates. Any emulation or replication of an entity or process can be considered multiple realizable, in that the replication instantiates the same state or event (however defined) as explained in detail in the beginning of the chapter in terms of the new approach to the mind–body problem being worked out in the current thesis. Although multiple realizability has been used to critique identity theory (e.g., Putnam, 1980), this can be called into question (Jaworski, 2011; Bickle, 2019). The anti-identity theorist multiple realizability argument may be presented as follows.

(1) Mental states are multiply realizable,
(2) Multiple realizable states cannot be identical (or reduced) to physical instantiations,
(3) Therefore, mental states are not physical substrate states.

Consider the mental state of pain. It can be said that pain is multiple realizable both within the same physical entity (each pain we experience is different from a previous pain experienced) as well as across species (an animal, or extra-terrestrial can be said to be in pain despite it having a different physical constitution than a human). Then, according to the argument, it stands to reason that pain cannot be identified with any particular physical substrate. For the functionalist, pain is

an organisational structure of causal functions abstractly understood that could be applied across species and substrates. As the functions can be described independently of the physical substrate that instantiates it, functionalism can be said to be non-reductive.

Churchland (1986) (see also Smart 2017 presented in sections 2.5.2.1), focuses on (2) in the above argument and counters this proposal by arguing that reduction (identity) in one substrate does not necessitate the denial of multiple realizability across substrates. Churchland discusses the phenomenon of temperature as a paradigmatic case of scientific reduction that is, nevertheless, multiple realizable. Temperature of gases is taken to be reducible to mean kinetic energy and is an identity claim (the temperature of gas *is* the mean kinetic energy of the gas molecules). However, temperature in plasma and temperature in empty space, although still temperature, is not reducible to molecular kinetic energy as the necessary molecules are not present in plasma or empty space. Furthermore, two separate gas volumes may present at the same temperature yet have different molecules, molecular movement (velocity, trajectories) and so on. Therefore, temperature has variation of reducibility relative to the domain of realization (e.g., it is reduced differently within gases to as it is in empty space) yet it is multiple realizable (it is still temperature). For Churchland (ibid.), a preferred understanding of reduction would then be that certain phenomena are reducible within various domains as well as multiple realizable across said domains.

The qualm that Churchland has with functionalism is not that functions may be used to describe causal properties, nor that functions are multiple realizable, but that functions are not a second order autonomous property (therefore, affirming Ll functions and denying Hl functions). Reductive Ll does not need to make a commitment as to which nomological level of function is necessary or sufficient and rather awaits further empirical (nomological) theories. As Churchland states, "...if human brains and electronic brains both enjoy a certain type of cognitive organisation, we may get two distinct, domain relative reductions. Or we may have one reductive account..." (Churchland, 1986, p.357). The question of which physical level is needed is, therefore, an open question although realizer functionalists, specific claim functionalists, and first order functionalists (Ll) may assume that lower physical levels are more necessary than presumed by the role-functionalists (Hl).

In essence, whether one holds to a non-reductive (Putnam, 1960, 1980; Fodor 1974, 1975) or a reductive stance (e.g., Churchland, 1986; Smart, 2017), multiple realizability is still feasible on both accounts. It, therefore, offers no metaphysical objection to the MUP, although would indicate variation in engineering and what levels are needed.

Assume that all the necessary and sufficient functions for the mind have been identified and are replicable. Would this functional isomorph be a mind or merely a mind simulation? This is the subject of three common functionalist objections that are discussed below in section 2.6.4. The first is that of Searle's (1980) Chinese room. The second is that of Block's (1980) Chinese nation argument, and the third is the philosophical zombie argument (Kripke, 1972; Chalmers 1996).

2.6.4 Objections to functionalism

In the Chinese room thought experiment (Searle, 1980) we are asked to imagine a man (who has no understanding of Chinese) in a room, who receives (unbeknown to he him) Chinese symbols. Once the squiggles (Chinese symbols) are in the room, the person references a manual that contains clear instructions what to do in relation to each symbol and writes down, according to following the manual, more squiggles that are pushed out the room. What is happening outside the room is that Chinese scientists are performing a Turing test and inserting texts (Chinese symbols) that have meaning, and what emerges from the room are answers that are meaningful. For Searle, the Chinese room thought experiment indicates that a person can manipulate symbols (syntax), while retaining no meaning (semantics), and this can be inferred from the man in the room performing the symbol manipulation without any understanding of the meanings. In this sense, syntax and semantics are separable and the computational theory of mind may well account for the syntax but will be unable to grasp semantics (meaning). For Searle (2004), meaning (semantics) is an essential component of the mind and, therefore, the implication is that any purely computation (representational symbol manipulation) will not be a mind.

There are many responses to the Chinese room thought experiment (see Preston and Bishop, 2002 for a collection of essays on this subject). The current thesis understands the thought experiment to rest on an intuition, namely, that the system (including the manual and the person manipulating the symbols) does not understand. If this is so, the question of whether symbol manipulation can lead to semantics is denied within the assumed intuition (that a symbol manipulation system cannot understand) and so appears to beg the question[31]. If an alternative intuition is upheld (i.e., that syntax may lead to semantics), then an intuition stalemate occurs as the assumption would be that the system does understand. For example, within cognitive neuroscience, the general un-

31 The alternative intuition (that syntax leads to semantics) is presented in Hofstadter's (1982b) thought experiment, where Einstein's mind is retained in a similar manual.

derstanding is that semantics is not found in any particular sub-system such as a neuron, or a particular lobe, but rather that the system as a whole (through the interaction of various unconscious parts), along with grounding in the environment, understands (e.g., Meteyard, *et al*., 2012; Kandel *et al*., 2013) through information processing (syntax). Furthermore, the sharp distinction between semantics and symbol manipulation is denied by some proponents of AI (e.g., Moravec, 1988; Minsky, 2006). Therefore, the assumption that symbol manipulation cannot lead to semantics is questionable, since it assumes an intuition whose opposite may also be assumed.

However, assume that the Chinese thought experiment is an effective rebuttal of the computational theory of mind. What would this mean for the MUP? For Searle (2004), in his biological naturalist view, the mind is a product, and is causally reducible to physical (e.g., neurophysiological) processes that include but extend beyond the information representational processes that occur in symbol manipulation. If all these biological processes were to be sufficiently replicated, then the result should be the same mind. Therefore, there is no significant metaphysical problem for mind uploading but rather an engineering one, namely, what components are required for construction (the B of the ABC options).

In Block's (1980) Chinese nation argument against functionalism, the neural functions of the brain are abstracted and performed by the nation of China. Each of the billion inhabitants is co-opted with a two-way radio to send signals to an artificial body. Each person behaves in a similar fashion to the inputs and outputs of a neuron. If this is designed to have the same functions as your brain "[i]t could be functionally equivalent to you for a short time, say an hour" (Block, 1980 p.276). The intuition that Block has, is that this system would not be conscious and, therefore, that functionalism is false. Aside from engineering concerns (e.g., can the nation of China be orchestrated to perform all the relevant functions at all the relevant neural levels[32]), it appears to lead to a similar intuition stalemate as Searle's Chinese room thought experiment. The functionalist would assume that if all the functions at all the relevant levels were replicated through some miracle of coordination, then the nation of China would, in fact, be conscious as this is the functionalist intuition (that isomorphic functions lead to the same mind). Therefore, Block's argument either can be denied on the basis of an alternative intuition, or on the basis of not considering the necessary functional granularity.

32 The two-way radio, at best, captures the on-off system of neurons firing but does not include the complexities of neural activities that processes at multiple levels simultaneously (e.g., Baars, 1988; Lycan, 1998).

If it is assumed that Block (ibid.) is correct, and that further physical processes that lead to consciousness are needed, then the MUP is in a similar to position as in responding to Searle's (1980) Chinese room. If the mind requires physical processes that are unique to biology (as is Block's position) then a biological substrate is needed for multiple realization. If Block is incorrect, then Hl-processing (as may be achieved through silicon based computers) may be sufficient. This, therefore, leads to Block being aligned with the B of the ABC options.

For both Block (1980) and Searle (2004), the mind is emphasised as a bio-mechanical process (Ll) where the 'over and above' aspects of the mind (that which is not achieved through Hl-processing) are the result of biological interactions. As stated above, if this is true, it would indicate the need for a biological artifact (option B) within the MUP. Although this may be true, it does not resolve the mind–body problem because the same difficulties emerge, namely, if the question is how low-level information processing could lead to emergent properties of the mind, then mechanistic processing would have this same difficulty as the information processing does. This was the difficulty that was pointed out in Leibniz's mill (1965). Leibniz imagined the mind as a mechanism and used the analogy of a mechanical mill. If we think of the nature of thoughts and imagine a mill performing that action, then, if we were to enter the mill, we would see all the physical mechanical interactions yet be none the wiser about the nature and emergence of thoughts. Both the mechanical and the information processing theories present us with the same problem: How can non-conscious (lower level physical processes) interactions lead to conscious results (higher level physical processes)? This is what Chalmers (1995) calls the hard problem and is discussed further in section 2.7.1.

A further argument that may be briefly presented here is that of philosophical zombies (Kripke, 1972, Chalmers 2014). Philosophical zombies are creatures that are identical in function (they behave the same, talk the same, write philosophical thesis the same), yet lack consciousness. From a third person objective perspective, there is no way to distinguish your zombie from you. The only difference is that you are conscious and the zombie is not.

The argument emphasises conceivability by asserting that 1) a philosophical zombie is in the first instance a conceivable entity, 2) that what is conceivable is possible, 3) that what is possible is metaphysically possible and 4) therefore, that zombies are metaphysically possible. The reliance on the conceivability argument has already been discussed (section 2.4.1) as a questionable approach to metaphysical assertion. For example, that an entity can be functionally the same but lack consciousness, can be inconceivable and incoherent to a functionalist (similar to a non-flying flight, a non-walking walk). Because, in the functionalist view, consciousness is itself a function and then

it is meaningless to be speaking of a being that is functionally the same but consciously not (a non-functioning function). Furthermore, the assertion of a metaphysical possibility, in some possible world (the philosophical zombie), does not necessitate that this is the case in our world (the nomological). If a zombie can be conceived of as resulting not from the standard physical functions then equally so a conscious being (a non-zombie) can be conceived of as resulting from said functions (the knife of conceivability cuts both ways). What can be asserted from intuition by one can be denied on the intuition of another (see Frankish 2007 for further intuition stale mates in relation to zombies). However, even those who assert philosophical zombies (e.g., Chalmers 1996) allow for the feasibility of mind uploading through functional isomorphs that result in consciousness (Chalmers, 2014; see section 2.7.1.).

2.6.5 Functionalism evaluation

Functionalism has been shown to be a diverse and complex perspective with multiple forms and presentations within the literature. The current thesis has demonstrated how many of these forms allow for the probability of the MUP, irrespective of whether the functions are external or internal (here including Ll, Hl , or a hybrid of both Ll and Hl functions replication). As long as functions are deemed to be descriptions of physical processes (whether those physical process be reductive standard physical properties and processes or further physical non-reductive properties and processes), they are likely to be replicable (the multiple realizable argument) and, therefore, the MUP is feasible if these conditions are upheld. The variations of functionalism have been shown to offer options for views based on a standard physical substrate or to include views based on further physical (in this section focusing on emergent properties) substrates. Nomological boundaries vary according to the functionalist view (e.g., a psycho-functionalist may assert that only the psychological processes are needing replication) and it is yet to be determined, as well as being beyond the scope of the current thesis, as to precisely where the boundary needs to be drawn. The following section now refers to property dualism which includes a further form of functionalism (Chalmer's section 2.7.1).

2.7 Property dualism (further physical properties)

This section now presents physical monist perspectives that assert that the mind has further physical properties (the mental) that are somehow distinct from standard physical properties (such as are commonly understood in physics). The emphasis here is on property dualism which is physical monist in that it asserts only one substance (the physical as broadly defined by the current thesis)

yet is dualist in asserting that the mental exists as a distinct property from standard physical properties. There is, therefore, one substance (the physical broadly defined) with two categories of properties (the standard physical properties and the further physical properties of the mental). The current thesis has focused on understanding of these further physical properties (the mental) as to exist either through emergence (the mental emerges as a new property based on interactions of standard physical properties and their processes at the lower level) or it may exist alongside standard physical properties (panpsychism).

2.7.1 Chalmer's panpsychic functionalism

Chalmers (2014) is a physical monist property dualist who has argued for the feasibility of mind uploading. Earlier, Chalmers (1995) put forward the 'hard problem' of consciousness asserting that the current functional analysis of the mind put forward by modern neuroscience (the easy problems) is insufficient to solve the problem of consciousness (the hard problem). The hard problem is the problem of subjective experience: "Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does" (Chalmers 1995, p. 3). Subjective experience, in this sense, is taken to be a brute fact of sorts that cannot be reduced to our current notions of the physical (standard physical properties). Rather, Chalmers (2015) hopes to integrate consciousness into the physical domain as a distinct property holding to a form of panpsychism. To distinguish between the identity theory reductions from Chalmer's property dualism (a version of non-reductive physicalism), a comparison may be made between temperature and electromagnetism, with temperature analogous to the identity theorist stance and electromagnetism analogous to Chalmer's stance. Temperature may be reduced to mean molecular energy where the term temperature is subsumed under a more refined scientific theory, whereas that of electromagnetism, where electric properties and magnetic properties are each retained, integrates these properties to form a new theory of electromagnetism. How this integration between the mental properties and physical properties can be achieved, according to Chalmers, is through producing a functional isomorph (sometimes termed by Chalmers as a functional invariant), which implies that should the fine-grained organisational causal functions be replicated in another substrate, then consciousness (the mind) would transfer to the new substrate. The fine-grained isomorphic functionalist may be seen as subscribing to a hybrid of Hl and Ll, where both the higher and lower level functions are necessary for emulation.

One argument Chalmers uses to explore how functions may relate to consciousness, is that of 'dancing qualia' (Chalmers 1996, Chap 7), where we are to imagine parts of the brain being

replaced by functional equivalent silicon chips from singular neurons to the entire system (here assumed the brain). First assume that absent qualia are possible, i.e., that a functional system such as in the proposed philosophical zombie arguments, where the functions are retained but the qualia (experience of seeing red for example) are not, is possible. What is likely to happen as each function neuron (or cluster of neurons) is replaced by the silicon functional equivalent? Either conscious experience will (1) gradually fade, (2) will suddenly disappear at some point, or (3) be maintained throughout. Between option (1) and (2), there is the possibility of consciousness fluctuating in and out of being and so the qualia may be said to 'dance' if this were to happen.

This thought experiment provides a way to validate functional isomorphs in retaining consciousness. Should they gradually fade (1) the artifact systems can be adapted until they retain consciousness or the replication of consciousness cannot be achieved (in which case, functionalism of this sort will be proven false). The same approach applies to (2) where the specific replaced artifact at the point of sudden qualia disappearance can be adjusted or disprove functionalism. The third option would validate the functional isomorph across substrates. Although Chalmers' functional isomorphs are taken to have a panpsychist orientation, the same would apply for strong emergence, where the emergent conscious property can develop at higher levels of physical functional replication. The dancing qualia thought experiment may further be applied to other forms of multiple realizable physicalism. For example, if one holds to illusionism (Ll → Hl), then the illusion of consciousness would equally gradually fade, disappear suddenly, or be retained.

2.7.2 Kripke's modal argument

Kripke's (1972) modal argument against type identity physicalists focuses on the common example of pain. The basic structure of the argument may be presented as follows:

(1) If pain is identical to C-fibre stimulation (or some other such neural correlate), then it is necessarily so (contingent identity is denied),
(2) It is not necessarily so (through the use of rigid designators pain is always pain – a felt qualitative experience – in all possible worlds),
(3) Therefore, pain is not identical to C-fibre stimulation.

Kripke's primary target is that of type physicalists (identity theorists) and he acknowledges that other physicalist/naturalist philosophers (here he mentions Nagel and Davidson) are likely "immune to much of the discussion" (Kripke 1972, p. 144). The modal method here relies on 'pain' as rigid designator (that 'pain' is pain in all possible worlds) and the possibility that pain

may be achieved without C-fibre firing (such as in other animal species, extra-terrestrials, or some other presentation). As such, Kripke's modal argument can be seen as an argument for multiple realizability.

There are many ways to critique Kripke's argument. For example, Jackson *et al*., (1982) argue against Kripke's modal logic argument by querying the use of rigid designators through exploring *de dicto* and *de re* readings. They assert that the *de dicto* reading of 'pain is necessarily pain' (there is no possible world where pain is not pain) is an empty phrase as pain requires contingent components such as the experience of pain and pain behaviour. Therefore, they and according to them, Kripke, must be considering pain as a *de re* reading that "waits on a theory of what pain is" (ibid., p. 216). According to Jackson *et al*. (ibid.), this theory of pain is likely to be a causal theory that may vary across species and so serve as a functionalist (a second order designation of a pain functions) type–type identity (the particular pain is related to a particular physical instantiation) theory. Furthermore, Kripke's argument relies on conceivability (1972), which has been called into question as a valid method for metaphysics earlier in this thesis (presented in section 2.4.1).

The complexity of these arguments fall beyond the scope of the current thesis but what is noted here, is that both Kripke (1972) and his critics mentioned here (Jackson *et al*., 1982) assert that pain (and, by implication, other mental states) is physically instantiated (even if of a different physical property traditionally described in science) and multiple realizable. Pain is not limited to the human species and pain and is, therefore, a general description of various contingent factors and pain is identifiable (although the strictness of identity may vary) with physical activities. For the MUP, this indicates that mental states may be achieved through various substrates and are, to some extent, substrate independent.

2.7.3 Davidson's anomalous monism

Davidson (2001) developed anomalous monism through emphasising the discrepancy (anomaly) between mental events and the "nomological net of physical theory" (ibid., p. 207). The general argument can be presented as follows:

(1) The interaction principle: Some mental events causally interact with some physical events.
(2) The principle of nomological character of causality (causation requires strict, deterministic, cause and effect laws).
(3) The anomalism of the mental (mental events do not hold to these laws).

These principles are all held by Davidson to be true. 1) He acknowledges the mental and physical distinction (manifest image) and assumes a causal relationship of some sorts. 2) He orientates his view of the physical closer to (and, by implication, his understanding of physicalism) what has been described in this thesis as a view based on standard physical properties and processes. The monistic assertion of Davidson is that while the mental is not describable in these strict physical terms, it is never the less dependent on the physical (supervenience). Davidson argues for a "version of the identity theory" (ibid, p. 209), while denying that mental descriptions can be reduced to physical events. This leads to token physicalism, in that each physical event (which is unrepeatable in Davidson's view) can be seen as a token with a more abstract mental description reflecting these various physical causes (token physicalism, see section 2.5.2.1). One of the reasons for the variation of descriptions between the mental and physical (as put forward by Davidson), is that the mental operates off principles of generosity (we assume others think and act as we do and employ explanations of their behaviour through this generous perspective) and the physical operates off laws of specificity (rigid laws that are required to be strict as in the physicists formulas).

For Davidson (ibid.) the mind–body problem can be categorised into four categories; nomological monism (that there are correlating laws that connect the mental to the physical), nomological dualism (that the mental and physical are distinct but are correlated in some way, and here Davidson includes parallelism, interactionalism, and epiphenomenalism), anomalous dualism (that they are separate substances but need not interact), and anomalous monism (that the descriptions of the mental and physical are irreducible but refer to the same physical causes). Because Davidson holds that only strict laws (the nomological character of causation) are the valid form of causation, the physical is always the cause and the mental (although it may present a causal explanation) is, at best, a partial description of these laws.

Although there are multiple avenues to critique and defend anomalous monism (see for example Hahn, 1999; Yalowitz, 2019), the current thesis emphasises the impact on the MUP. Because the causes of the mental, within anomalous monism, are the standard physical events, these causes should be replicable (the multiple realizable physicalist argument). As there are only physical causes, the disparity between the mental and the physical is one of disparate description and, therefore, the descriptions of the mental (our current understanding and experience of mental events) will be retained should the same physical causes be replicated. Put within the previous depiction of these relations in terms of lower level and higher level processes, anomalous monism can be simplified to Ll → Hl, where replication of Ll would be sufficient for the replication of Hl.

2.7.4 Strong emergence and property dualism

Emergence here is distinguished from panpsychism because the panpsychist asserts that the mental does not emerge but is always present (i.e., the mental is a further physical property that runs parallel to standard physical properties), whereas in emergence theory the mental property emerges from standard physical properties and processes. In this sense, Davidson (see previous section) may hold to strong emergence (the supervenient mental properties are caused by the standard physical properties and processes) and Chalmers (see above) is a panpsychist. Both may be defined as property dualism as both assert mental properties as something further than the standard physical properties, although Davidson's emphasis on descriptive disparity between the mental and the physical actually leaves his anomalous monism in a category of its own. As stated earlier (section 1.5.6), emergence can be broadly categorised into strong and weak emergence, where weak emergence is causally reductive and strong emergence is causally non-reductive (Stephan, 1999). Strong emergence, therefore, implies a form of property dualism because mental properties are held to be independent (to some extent) from the lower order properties (generally described in standard physicalist terms and processes).

Both weak and strong emergence base the emergent property (here meaning the mental) in standard physical properties and processes. The standard physical properties interact at the lower levels and a new property emerges and, therefore, reflects the current thesis's addendum to Sellars' remarks on things and how they hang together to include things interacting to form new things (e.g., atoms interact to form molecules). Therefore, replication of these standard physical properties and processes will result in the same mental properties irrespective of whether a weak or strong emergent stance is asserted. Furthermore, both weak and strong emergence may assert that the mental is causal (e.g., Ll ←→Hl). An example of an emergent property being causal may be seen in the trajectory of a dust particle within a liquid or solid substance. The solidity or liquidity in which a dust particle is embedded will influence the trajectory of the particle as it moves downhill and, therefore, the higher order property (liquidity or solidity) causes different trajectories. The property of solidity or liquidity is an emergent property from activities at the molecular nomological level (it is caused by these activities), but this property (solidity of liquidity) is also itself in a causal relation to the lower levels (i.e. the trajectory of the atoms within the dust particle). The primary emphasis lies between the difference between autonomous causation and reducibility with the strong emergent view asserting that mental properties are autonomous non-reducible properties, whereas the weak emergent view asserts that the mental is non-autonomous and is reducible.

The strong emergent view is, therefore, akin to Hl functionalism (e.g., second order functionalism, functional state identity theory, and so on), whereas the weak emergent view is akin to Ll functionalism (first order functionalism, functional specification claim, and so on) discussed in section 2.6.1. One of the difficulties facing the strong emergent view is that there is no uncontested example outside of referring to the mind that has been used to illustrate strong emergent properties (Seager, 2007), such as the emergent property of solidity being nevertheless reducible and non-autonomous (weak emergence).

Assume that strong emergence is true. The higher level property of the mental would then be autonomous and non-reducible. First, because the strong emergent property emerges from standard physical properties and their processes, the same strong emergent property (e.g., the mental) would emerge if those properties and processes were replicated. Second, because physical monism is upheld in this view, the properties would, based on the multiple realizability argument of the current thesis, be multiple realizable and, therefore, likely replicable in an artifact. The current thesis has emphasised that nomological boundaries may be drawn in relation to artifact creation (the example in section 1.5.6 of a carpenter making a table of wood without having knowledge of atomic properties and processes that are present in the wood was used to illustrate this). Because the strong emergent causal properties and processes occur autonomously at the nomological level of psychology, then this nomological level is all that is needed for replication.

However one understands physical emergence, it is congruent with the MUP. The physical monist underpinning of property dualism allows for the possibility of multiple realizability because any physical process can, in principle, be replicated. Any multiple realizable system requires sufficient causal knowledge and sufficient knowledge of the physical entities that interact to have the same function. Once this is known (which is not the case at present), then it is a matter of acquiring the knowledge of the system and the components needed to instantiate the system (which is an engineering problem).

2.8 Summary

The current chapter has explored some of the options that have been presented within the philosophical literature with the aim of identifying their potential relation to the MUP. The emphasis has been on the mind–body problem and what this means for replication of a particular mind. The thesis has presented multiple realizable physicalism as a perspective that integrates multiple physicalist perspectives and that allows for the metaphysical feasibility of the MUP. Furthermore, it has been shown that multiple realizable physicalism is a necessary and sufficient condition for replication

of a particular mind, although the thesis has not established what precise form this multiple realizability necessarily should take and has worked with a broad interpretation of the concept as explained in the beginning of the chapter (section 2.2.).

Section 2.4 addressed the possibility that the mind is based in a substrate that is not physical (here focusing on substance dualism). As the MUP aims to create an artifact that may instantiate a particular mind, the artifact would need to be within the realm of human engineering, which is the field of physical objects and their creation. Therefore, if a mind were instantiated in non-physical substrates, it would make the creation of a physical artifact to instantiate the mind not feasible.

The thesis then turned to discussion of physicalism (in the broadest sense), which was further categorised into views that have a standard physical substrate basis or, alternatively, may have a further physical substrate basis. Beginning with views based on standard physical substrates, it was shown that these emphasise that the mind is a process rather than an entity, with variation in external processes (e.g., behaviourism) or internal processes (e.g., identity theorists). If the mind is identified with processes of standard physical substrates (internal or external), then a replication of these processes should result in the same mind.

The internal processes can, in effect, be further categorised into higher (Hl) and lower (Ll) levels of analysis with the higher being associated with the mind (psychology) and the lower the body processes (biology), which reflects the primary distinction of the mind–body problem. This leaves certain options available. First, the Ll = Hl, in which case, the mind is identified (with various levels of granularity) to the physical processes. In this perspective, the mind–body problem is resolved through equating the two features (features of the mind and features of the body) as only appearing (manifest) to be disparate but in reality (metaphysically) being the same. Second, the option of the lower level leading to higher level effects that are not causal (Ll → Hl). This does not necessitate that the mind is not an ontological entity. For example, a rainbow (Hl manifest appearance) may be said to be a non-causal illusion, but it could fall under a person's ontology (i.e., the belief that rainbows exist). Despite being non-causal, the rainbow is still a physical event and is based purely on Ll activities (e.g., light refraction) that are replicable (to replicate a rainbow one simply needs a prism). If any of these options are the case (Ll = Hl[33], or Ll → Hl), then replicating the Ll at a sufficient level of granularity will result in the same mind.

33 The possibility of Ll = Hl being non-replicable through the assertion that the mental type of a specific mind being identified only with a particular substrate (flat realizer), was considered. If this were to be accepted, the specific mind would not be multiple realizable and the MUP would not be feasible. However, the current thesis has emphasised type

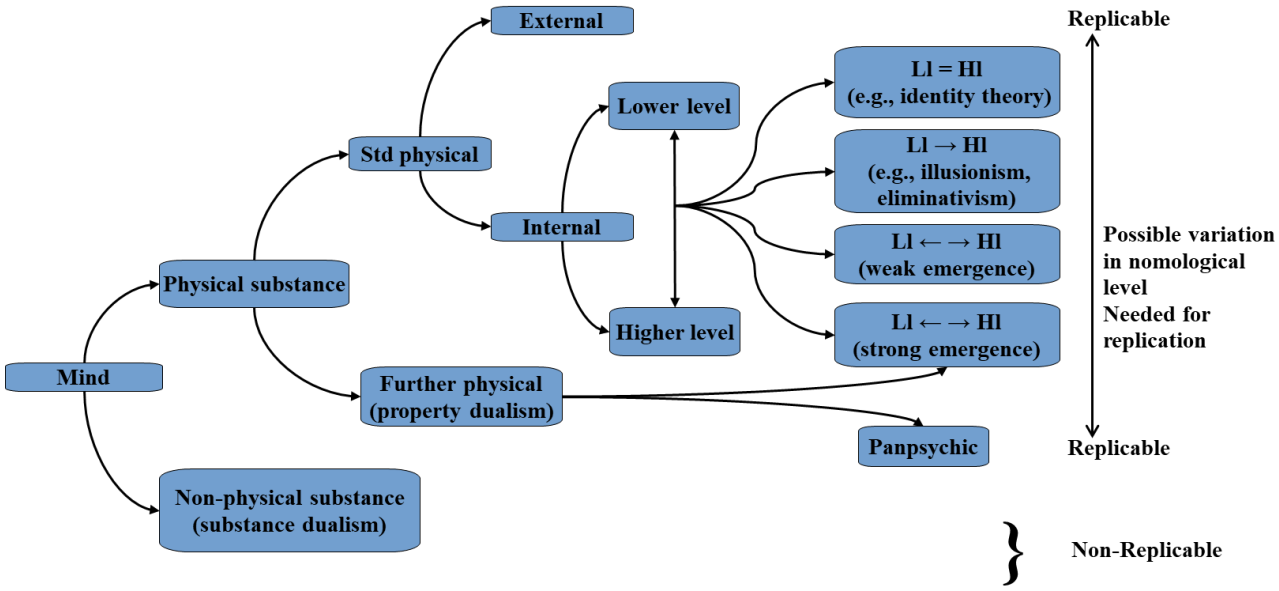

Figure 2-6: Summary of replication options in relation to mind–body problem

The thesis then turned to functionalism, which acts as bridge between the standard physical substrates and further physical substrates. The emphasis for functionalism was on causal processes (which distinguishes a functional system from all and any process that may occur). Furthermore, functionalism asserts that the mind is what the brain does (or at least the causal aspects), which is a metaphysical claim. Because the mind is causal (the definition of function), functionalism ignores Ll → Hl perspectives. Functionalism can then assert Ll = Hl, in which case, many of the functionalists (Lewis, 1966 Armstrong, 1968, Place, 1956, Smart 1959, 2017) are also identity theorists. The same inference for the MUP may, therefore, be applied in replicating Ll to replicate Hl. However, functionalism also may be seen as an emergentist view (e.g. Putnam 1960; Oppenheim & Putnam, 1958) where the Ll processes lead to emergent features. If emergence is held, the Hl can be said to have causal influence in its own right and so the arrow of causation moves in both directions Ll ← → Hl. This emergence may be deemed as weak emergence (reductive physicalism) or as strong emergence (non-reductive physicalism). If Hl are causal in their own right, in order to replicate a mind should the Ll features be incorporated or would the Hl replication be sufficient (e.g. the higher order property of solidity is needed to replicate a table without the need to understand or be involved with the atomic level that leads to the solid property)? And so, a question of

identity theorists who allow for multiple realizability as defined by the current thesis and this form of type identity is what is being referred to in the above section.

where to define and apply the nomological boundary was raised. It may either be only necessary to include the Hl (e.g., representations) or it may also require some of the Ll (e.g., unconscious neural functions).

The possibility of strong emergence allows for new mental properties (that are distinct from standard physical properties) that are non-reductive to emerge. This is the one option for those who hold to some form of further physicalism (still maintaining a physical monism, while asserting that there are mental properties) which has been considered in the current thesis under a discussion of property dualism. Aside from properties emerging from Ll standard physical properties and processes, there is the option within property dualism that these mental properties co-occur at all physical levels (panpsychism). Property dualism (of both sorts) allows for the possibility of the MUP, in that although its advocates argue for the mental being a further physical property, they argue that it is physical (with a broadened interpretation) nonetheless (therefore, being a physical monism). As physical properties (as we currently understand them) are replicable, it may be inferred that any further physical property would also be replicable (see the argument for multiple realizable physicalism in section 2.2).

In both the standard physical and further physical substrates, it is has been shown that replication of a mind is a possibility (and, therefore, metaphysical) so long as the view adheres to multiple realizable physicalism. The variation of physical views, rather than implying a metaphysical difficulty, imply variation of emphasis of what nomological level should be replicated. If one holds to the view of the mind as external processes, then it is these processes (the interaction of the entity with its environment) that need to be replicated. If one holds to the view of the mind as internal processes, then either the Hl or the Ll, or a hybrid of both, would need replication. If one holds to a form of property dualism (whether this is achieved through strong emergence or panpsychism), the replication needs to occur only on these properties and features (what is defined here as the Hl).

The thesis has argued for the metaphysical feasibility of replicating a particular human mind and asserted that many of the physical monist views are congruent with this aim so long as they also hold to multiple realizability (multiple realizable physicalism). The thesis now turns to the question of personal identity and the persistence problem. In essence, if it is conceded that a particular mind is replicable, would this replication be the same person?

3 The persistence problem and the MUP

3.1 Introduction

Previously, this thesis has addressed solutions to the mind–body problem as they relate to the MUP. In essence, what was evaluated was what the nature of a mind is and how this relates to the feasibility of replicating a particular mind. It was demonstrated that multiple realizable physicalism may act as a unifying concept across physical stances that would allow for the metaphysical feasibility of the MUP. Furthermore, the previous chapter asserted that the mind was more likely a process as opposed to an entity. Within this chapter, the thesis thus assumes a multiple realizable physicalist stance (with an emphasis on mind as process) and, therefore, it is inferred that the mind is replicable. From this assumption, the current chapter focuses on the persistence problem as it relates to personal identity (i.e., what it means for a person to continue over time) as this problem relates to the MUP.

Consider an artifact is created, through whatever means, that results in a sufficient replication of your mind (as was argued for in the previous chapter). The artifact sits across the room from you. Given that the artifact has the same mind as you have, the artifact thinks and feels as you do; after all, it is a sufficient replication by definition. Now consider the question of whether the replicated artifact *is* you. Are there now two you's? If not, who is the real you? The questions raised by this scenario (and many such thought experiments, which will be evaluated in section 3.4) demonstrates that it may be one thing to replicate a mind and another to assert that this replication maintains the same personal identity as that of the 'original' mind. In essence, the chapter asks if replication of the mind (as is the agenda of the MUP which includes all properties and processes) would result in the same person and, therefore, this chapter evaluates what the metaphysical nature of you, the self (personal identity) that continues over time (the persistence problem), really could be.

There have been multiple solutions posed in an attempt to resolve the persistence problem of persons (e.g., Anscombe, 1975, Perry, 2008a, 2008b; Olson, 2017; Gallois, 2016), with the two dominant categories that have been established as that of the biological (e.g., Williams, 1973; Olson, 2007) and that of the psychological (e.g., Parfit, 1984; Lewis, 1987). Although "philosophical orthodoxy has assumed personal identity is grounded in psychological facts" (Olson, 2007, p. 135), this should not be simply adopted, as this would be an argument from authority. As with the previous chapter, the MUP serves as a lens from which the traditional literature can be evaluated and aspects of the tradition emphasised for relevance to the MUP. In this chapter it will be argued that

the psychological solution is the preferred metaphysical stance that allows for the MUP to be feasible.

Section 3.2 defines the persistence problem and argues that the basis of the problem is the need to acknowledge that persons (however defined) are dynamic systems (however defined). The emphasis on continuity as a necessary and sufficient condition for persistence of personal identity is presented as the starting point for potential solutions to the problem of persistence of identity. The need for partial psychological continuity (as opposed to absolute continuity) is put forward as the necessary and sufficient condition for persistence of personal identity.

Section 3.3 orientates the chapter through situating the philosophical discussions within the needs of the MUP, which emphasises artifact design of a substrate and its processes. This orientation provides a novel stance from which to examine the persistence problem. The novel stance is in contrasting partial and absolute continuity alongside absolute and partial identity, respectively. It will be argued that absolute identity and continuity are not feasible options and, therefore, that partial identity and continuity are preferred. In relation to substrates and processes, the biological position of personal identity is aligned in this thesis with focus on continuity of the substrate, and the psychological position is aligned with focus on continuity of the processes instantiated in the substrate.

If a person is identified with the substrate (the person as a particular numeric biological entity), then the MUP is not feasible as the artifact of the MUP would not be 'you'. If, however the person is identified with the psychology, then so long as the artifact instantiates that same psychology (however defined) the 'upload' can be said to instantiate[34] the same person. In this sense, the psychological view may be formulated as the mind = person and what holds true for the mind would hold true for the person through the principle of transitivity (i.e., minds = persons, minds are multiple realizable and, therefore, persons are multiple realizable).

Section 3.4 evaluates various thought experiments and what can be inferred from these in relation to persistence of personal identity. The thesis evaluates the biological and psychological

34 The reader is reminded that the term 'instantiate' (as discussed in section 2.2) is broadly defined in the current thesis and includes instantiations at multiple nomological levels (e.g., atoms interact to instantiate molecules). For traditional philosophy of mind, it may be more common to use the term in the sense that the mind is instantiated in the body whereas 'emergence' may be used in the sense of the mind emerging from the body. As the current thesis's broad use of 'instantiate' (also see section 2.2) may include both (e.g., that a mind emerges from a body may be said to be an instantiation similar to molecules emerging from atoms and their interactions), it is used as such within the current chapter (i.e., an instantiation of a mind/person is an instance of the phenomenon mind/person that may be both instanced in or emerges from a substrate).

options as they relate to identity and continuity of process (psychology) and entity (biology), with biological continuity of the substrate in space and time being emphasised as an area of distinction. This is because the psychological option, if upheld, would reflect the work of the previous chapter (which demonstrated how a mind is readily congruent with the MUP), whereas the biological option, if upheld, would negate the MUP. Because this section addresses thought experiments, the aim is to excavate the intuitions that are related to the persistence of personal identity, with the further sections of this chapter further evaluating those intuitions through philosophical enquiry.

Section 3.5 queries whether personal identity is best understood as a numeric identity or a qualitative identity and focuses on, to borrow a phrase from Parfit (1984, p. 215), "what matters" in relation to the persistence of persons over time. This section argues against numeric identity as being relevant to the MUP as the MUP relates to persons as minds and deems particular substrates as non-necessary. Qualitative identity (or some variation thereof) allows for some conditions for persistence of identity to continue as necessary, while allowing for other conditions to be discontinuous (the nature of partial continuity). It is argued that qualitative identity is the preferred stance for personal identity and that psychological conditions are necessary and sufficient for continuity of persons.

Section 3.6 develops the type–token distinction in relation to the MUP and the persistence problem (as opposed to the distinction as it relates to the mind–body problem). For the biological solution to hold, the person type needs to be identified with the particular substrate (the token). For the psychological solution, the token (a psychological process instantiated by a substrate) is not identified with the person type, as the type may be instantiated across tokens. If the person type can be instantiated in an alternative substrate (token), the 'upload' can be said to be the same person (as type is identified with the person). Therefore, for the MUP to be feasible, the person type cannot *be* (identity claim) the token that instantiates the type.

Section 3.7 evaluates the 'branching' and 'non-branching' concepts as they relate to the MUP. It is argued here that based on the previous sections (that the person type may be instantiated across tokens and that qualitative, psychological, partial identity is preferred), a conclusion that allows for branching is the most probable outcome.

As with the previous chapter, the subject matter is too broad and complex to attempt to resolve all the various disputes within the field. Therefore, the aim of the current chapter is to identify various areas of dispute as they relate to the MUP, as well as potential options for resolution.

3.2 The persistence problem and the need for continuity

In this section, it is argued that the persistence problem's core dilemma is how the claim that the identity of a phenomenon (whether defined as an entity or a process), can be retained over time, given that the phenomenon is dynamic (it changes over time). In essence, how can something that changes be the same? Although themes within this section are evaluated in greater detail throughout the chapter, they are introduced at the outset to clarify the nature of the problem under discussion.

The current chapter focuses on the human person as opposed to the broader concept of all possible types of persons (e.g., extra-terrestrial persons, artificial intelligence persons), yet broadens the notion to include what a human person, which is currently instantiated in a biological substrate, may adapt to become. The question is, therefore, not limited to "what we are" (Olson, 2007), but includes what we may become. However, this does not negate the question of what we are at present, as this may influence the possibilities of what we may become. For example, if we are in essence biological animals, then by definition we could not become a non-biological artifact. However, if we are in essence information processes (currently instantiated in a biological substrate but not limited to this), then we could indeed become a non-biological artifact.

In order for the MUP to be successful, it requires that the person (however defined) continues to exist within a substrate alternative (whether a biological or non-biological substrate) to the original substrate (the original biological substrate). This raises metaphysical questions as to what the nature of persons is, how are persons identified, what is required for survival across substrates, and how is it that a person persists over time. Although these questions could be addressed individually, the current thesis opts to subsume them under the primary problem of persistence of personal identity.

For a person to persist, it is assumed that persons (however defined) continue over time and, therefore, that persons are in some sense temporally continuous. This continuity may relate to mental events (e.g., Lewis, 1987; Campbell, 2006), resulting in the psychological solution to the problem of persistence of identity through time, or to the continuity of the biological organism (e.g., Wiggins, 1967, Snowdon, 2014), resulting in the biological solution to the problem. The problem of persistence arises because persons are also continuously changing. We do not retain the same biological matter over time (e.g., cells and atoms are continuously replaced) and we do not retain the same psychology over time (e.g., past cognitions are different from current cognitions which will be different from future cognitions). Change is, therefore, a necessary condition of per-

sonal identity and reflections on the persistence problem attempt to find resolutions to how a changing (dynamic) person can still retain the same personal identity. There is, therefore, continuity of persons despite continuous change. The problem may be formulated as follows:

> If person x exists at t1 and person y exists at t2, what are the conditions under which person x is person y?

The current thesis defines continuity along a continuum from absolute continuity (that which continues is precisely the same in every conceivable way) to partial continuity (that some conditions for identity continue such that an identity claim may still be made). Continuity in relation to personal identity is defined as the collection of necessary and sufficient conditions under which an identity claim may be made. A similar continuum is presented later in section 3.5 as it relates to absolute identity and partial identity. The emphasis in the current section is simply to note that the argument of the thesis will be, that for persons to continue, partial continuity is preferred. The idea of absolute continuity is incongruent with personal persistence because if there is absolute continuity there can be no change. Therefore, partial continuity (continuity of some conditions for identity) must be adopted. Partial continuity of some sort is also the defining condition of solutions presented to the problem. For instance, in the context of the biological solution, psychological conditions are denied as being as a necessary and sufficient condition for the continuity of persons and in the context of the psychological solution, biological conditions (e.g., the continuity of a particular body substrate) are denied as a necessary and sufficient condition for the continuity of persons.

Because continuity, of whatever kind, is the necessary and sufficient condition for persistence of persons, the question arises as to what conditions are deemed necessary and sufficient in their turn, to ensure continuity (without these conditions the continuity of identity would cease), and what are deemed non-necessary and non-sufficient (if these conditions are present at one time but not another personal identity may yet persist). This distinction could be established as the traditional view of what are essential conditions and accidental conditions for identity (Gallois, 2016). It should be noted that, if such a distinction (essential and accidental) were to be emphasised, it does not necessarily commit the thesis to a substance/phase sortal distinction that aligns essential conditions to substance sortals and accidental conditions to phase sortals (see section 1.5.4).

Being ‘essential’ is more loosely defined in the current thesis in terms of necessary and sufficient conditions, i.e., what matters, for a phenomenon to exist, irrespective of whether this phenomenon is defined as a substance or process. For example, a wave function is ‘essentially’ a process of physical entities interacting (consider the slinky mentioned in the previous chapter that

transitions from a plastic coil to a metal coil, while retaining the same wave function). It is, therefore, not 'essential' that certain processes retain the same substance, or substrate for that matter, to be considered an 'essential' condition. A process may, therefore, be 'essential' but not related to a particular substance.

In essence, the current thesis is asking what matters in the persistence problems and allows for multiple interpretations and stances. In its simplest form, what matters is that which, if removed (whether entity or process), would disallow for the continuity of persons. The term 'essential' is used as such (interchangeable with 'what matters' and 'necessary and sufficient conditions') from here on within the current thesis. Although the concept of identity is further developed in section 3.5, the current section relates the concept of personal identity to the persistence (continuity) problem, in that both relate to the continuity of a phenomenon and how this phenomenon is identified as the same phenomenon through time. The current section does not elaborate on what these continuity conditions are, that being further developed throughout the chapter, but establishes the persistence problem as:

(1) Personal identity relates to the phenomenon of a person that can be said to be the same over time.
(2) Personal identity relates to the phenomenon of a person that can be said to change over time.
(3) Therefore, either personal identity is contradictory and, therefore, arbitrarily assigned, or it may be defined as the continuity of certain conditions (partial continuity).

To resolve the persistence problem, identity claims can either be arbitrarily assigned or based on intrinsic properties or processes. In the arbitrary assignment, there is nothing about the phenomenon's existence that does not rely on extrinsic attribution of identity. Existence of the subject is, therefore, dependent on convention. Consider money in the form of a bank note. There is nothing intrinsic to the piece of paper that gives the note the identity and function of money. It could be equally a different object (e.g., a plastic square) and retain the same identity of money. The concept of money is, therefore, entirely dependent on the continuous convention of society in determining what it is. Without the ongoing convention from extrinsic processes, the function and identity of the phenomenon ceases. This is not to say that this form of arbitrary extrinsic identity is meaningless, after all money is a common staple and function within our everyday lives, yet it is arbitrary (in the sense described here) nonetheless. Now consider a cat. There may be a speech

convention that allows us to represent the creature with different names (e.g., 'kat' in Afrikaans), but if there were no language to name the creature, the creature would still exist. For example, dinosaurs existed long before we named them and existed without any such convention. One arbitrary solution to the persistence problem is to assert that there is no phenomenon that is 'you' (Russell, 1919; Ayers, 1936/1962). For example, the designator 'you' may be an arbitrary speech convention such as the 'it' in 'it is raining' (Anscombe, 1975, p. 55). The persistence problem is, therefore, resolved, in that there is no intrinsic 'you' that needs to persist, only a convention that persists.

A further such view of arbitrary assignment could be Nozick's (1981) closest continuer argument. Nozick considers the case of the Vienna circle (a club or society of philosophers and scientists during the 1930's), where the members left Vienna, some going to Istanbul and some going to America, both continuing the circle's programme and retaining the name 'The Vienna Circle'. Which is 'the' Vienna circle? Nozick, assuming that there can only be one Vienna circle, asserts that the identity continues with the one which is the nearest/closest continuer. This assumption, that identity cannot branch (non-branching), is addressed later in section 3.7.

When Nozick turns to persons "we are not willing to think that whether something is *us* can be a matter of (somewhat arbitrary) decision or stipulation" (ibid, p. 34); yet this is what Nozick does in elaborating on what may be best determined as "closeness". For example, Nozick presents various thought experiments such as a) a perfect duplicate of you is made and you remain alongside the duplicate. In this instance, the duplicate is deemed by Nozick not to be you as the original substrate is the closest continuer (i.e., there is greater/closer continuity of body). b) A duplicate is made and your original body expires. In this scenario, in Nozick's view, the duplicate is the closest continuer as your original body is no more and, therefore, the closest continuer is the duplicate, 'you' are the duplicate. The idea that in one instance the duplicate is asserted to be the true you and in the other it is not, and that this attribution is dependent on activities outside of the duplicate's intrinsic conditions (e.g., whether the original body dies or not), leads the current thesis to view the closest continuer as a close continuation of the arbitrary extrinsic designation of personal identity. Furthermore, it appears to the current thesis, that the problem or the closest continuer rests on the assumption that there cannot be two persons in any scenario that have an equal claim to designation 'you' (i.e., that 'you' is essentially a numeric identity that is contested by the current thesis in section 3.5).

The assertion that personal identity is based on arbitrary assignment from extrinsic conditions (as necessary and sufficient conditions) raises further difficulties. First, consider that if true, and identity is a purely social construct, then the proponent of such a view would need to accept that when such assertions in the past were made (e.g., within colonial times non-European races

were deemed animals as opposed to persons), that these assertions that deny person status would reflect 'the truth of the matter' (i.e., that the attribution of animal status by external arbitrary conditions would in fact determine that non-European races were metaphysically non-persons). Second, consider the thought experiment where every social interaction were (for some unknown reason) to suddenly assert that you were no longer a person. Would this be sufficient for you to state, "how about that, I'm not a person after all, everyone else has said and, therefore, it must be so". It is unlikely that one would assert this and so, if arbitrary assignment is denied, the continuation must be based on intrinsic conditions. However, if this view of extrinsic arbitrary assignment of identity were upheld, then all the MUP would need to do to ensure personal continuity would be to locate the upload within an upload accepting community, such as a transhumanist community (see for example Braidotti (2013) for a social constructionist approach to transhumanism). If arbitrary extrinsic conditions are rejected, then intrinsic conditions would need to be established as necessary and sufficient to personal identity.

Arbitrary attribution is not the approach to be addressed within the current thesis, which assumes that 'you' are not arbitrarily assigned, but rather are maintained through intrinsic properties or processes. However, intrinsic conditions cannot imply absolute continuity (or the second premise of persons changing over time would be denied) and so partial continuity must be accepted. The question then is, what are the necessary and sufficient intrinsic conditions of personal identity such as to allow for change as well as continuity of identity. This emphasis on intrinsic conditions is the primary focus of the current chapter, with further emphasis on the two dominant stances of biological and psychological continuity, with each arguing for their respective stances as to what conditions (biological or psychological) are the necessary and sufficient conditions at issue.

3.3 Substrates and processes (biological and psychological solutions relating to the MUP)

First, this section introduces the traditional biological and psychological solutions to the persistence problem, which may be defined as follows:

> "The Psychological View of Personal Identity: X at t_1 is the same person as Y at t_2 just in case X is uniquely psychologically continuous with Y" (Shoemaker and Tobia, 2019, p. 7).

> "The Biological View (aka animalism): If X is a person at t_1, and Y exists at any other time, then X = Y if and only if Y's biological organism is continuous with X's biological organism" (Shoemaker and Tobia 2019, p. 9).

The definition of the psychological solution, as stated above, which includes reference to 'uniqueness', is queried in later sections because, within the MUP, persons are assumed to be replicable. How could that which is replicable be unique? What is the basis of this uniqueness and in what sense, if any, can it be retained within the MUP? Uniqueness, therefore, needs to be further clarified and cannot be assumed as an *a priori* element of the metaphysics of the MUP. The nature and necessity of uniqueness is developed within the current chapter in section 3.5, where numeric identity is defined as a one to one correlation, therefore indicating uniqueness; within section 3.6 on type–tokens, where the token only (or type = token) view is that persons are only a particular unique token; and section 3.7, where non-branching is the view that there can be only one unique branch of the personal identity tree. The current thesis at this junction thus removes uniqueness from the psychological definition.

The current thesis orientates the two solutions of biology and psychology to the persistence problem within the MUP context, through the use of terms such as substrates and substrate processes (substrates and what the substrate does). The substrate in this chapter is defined in terms of the body of a person (the biology), whereas the processes are aligned with the psychology of a person (based on arguments from the previous chapter). In the context of the mind–body problem, the mind was (in the previous chapter) identified with what the substrate does (whether these be internal or external processes) and this stance (that the mind is what the substrate does) is continued in the current chapter. The substrate is, therefore, the current physical (however defined) entity that produces/performs/instantiates the processes that are related to persons. From the physicalist perspective, both the substrates (the physical thing) and what the substrate does (the physical processes) are physical phenomena. It is accepted by the current thesis that substrates have properties that allow for processes to be instantiated (e.g., a wave function can be instantiated with a slinky but not a singular brick as these substrates have different properties). The substrates that are put forward as potentially instantiating the same person are, therefore, complex artifacts with appropriate properties (e.g., a biological clone or a synthetic computer of sorts may possibly instantiate a person but a brick would not). In the context of substrates for the biological solution the argument, therefore, is not that a substrate of sorts is needed (as that is the physicalist view of the MUP), but rather that it is the particular substrate (the current biological body) that matters for persistence of persons. When talking of the substrate-only option throughout this chapter, it is within the current thesis intended to mean the particular substrate (the body) and, therefore, such talk indicates the

biological solution. At this junction it is not predetermined whether it is the substrate that is essential to the person, the process that is essential, or both that are essential, but to simply to affirm that persons are related to things (entities) that do things (processes). Furthermore, this emphasis on substrates and processes allows for continuity with the previous chapter, which will be used to integrate aspects from the mind–body problem and the persistence problem, as presented through the exploration of the MUP (to be presented in chapter 4).

Evaluating persons in relation to processes and things is not new to the philosophy of mind literature. For example, Perry (2008b) compares personal identity to a baseball game stating: "I believe this general framework should be applied to the concept of a person" (ibid., p. 10). He talks of how a unity relation may be attributed to a game that has different events across different times yet be the same game. However, unlike the current thesis, Perry assumes that "after all, baseball games are not 'things' in the ordinary sense, but 'processes' … persons are not processes ... but ... we can think of his life or personal history as a process" (ibid., p. 10).

Perry makes no attempt here to validate his claim that persons are things and not processes[35]. This chapter queries the assumption that a person is a thing and asserts that this assumption presents the same category mistake identified in the previous chapter of conflating or separating the mind with the body based on entities (confusing things with how things hang together). Rather, it is argued here, that both minds and persons are processes (not the substrate that instantiates these processes) and that, through this stance, both the mind–body problem and the persistence problem may be better understood and, furthermore, would allow for the feasibility of the MUP.

The current thesis introduces a distinction here (to be developed further in section 3.5) between a space-time instantiation of a process and space-time continuity of the substrate. All physical processes at the nomological level of daily life (here ignoring quantum levels and so on) require physical properties in space-time in order for the process to occur (space-time instantiations). However, physical processes are not necessarily bound to the initial substrate (e.g. Perry's baseball game that can change locations, players, and yet retain the identity of the same game). Consider a domino cascade where one domino knocks the next domino, which in turn knocks the next domino, and so

35 Elsewhere, Perry (e.g., 1976) puts forward the view that, although psychological continuity is preferred, this continuity is nevertheless bound to the biological substrate (PI = Substrate + Processes – see Figure 3-1). Furthermore, his notion of psychological continuity is numeric and is a 'further fact' on top of the processes performed by the substrate (this is a version of the 'simple view' of personal identity rejected by Parfit's, 1984 and 2016, 'complex view'). As will be discussed further in sections 3.5 and 3.6, when a person is identified as a unique (non-replicable) thing/entity, the claim in this thesis is that there is a category mistake being made.

on. Now consider that half way through the cascade, the second half of the dominoes are moved such that the last domino of the first half falls without contacting the first domino of the second half section and, therefore, the cascade stops. Later, someone re-instantiates the cascade from the first domino of the second half section and so the cascade continues to the last domino. Now, is this one domino cascade or two? Does it even matter if we call the first half cascade A and the second half cascade B? What has happened, is that there has been a breach of temporal (cessation and activation) continuity and spatial continuity (the second section of dominoes moved location), yet it is at the least feasible to say that it is the same process.

Now suppose that the second half of dominoes is removed (prior to the last domino of the first half falling) and a different set of dominoes is placed in its place, so that as the last domino of the first half falls, it contacts the first half of the new domino half. Now there is continuity of processes that occurs across substrates (the first domino set up had numeric parts that have been replaced by alternative parts but the process continues) without breach of time (there is no cessation within the process) and space (the new domino parts occupy the same space location as the original parts did). Assume that the original dominoes are made of plastic and the setup is 10 metres long and, as the cascade is progressing, someone places more dominoes (made from metal with similar enough properties – e.g., weight, size – to allow for the continuity of the process) in sequence such that the cascade continues for 20 metres. Is it the same cascade? It has two different substrates, it has been developed after the initial substrate was arranged, and so on. Now suppose the domino cascade at some point diverges into two separate cascades (branching), a second line of dominoes is continued along two distinct paths. Is this the same cascade, or are there two separate cascades (they may vary in pattern from the initial point of divergence) that have the same cascade origin?

The domino cascade example illustrates that continuity of processes may be coherently defined as an extension of a previous process that, in principle, a) need not require the same substrate/substance, b) may cease and be restarted (temporal discontinuity), c) may be moved spatially (spatial discontinuity), and d) may continue from a singular process to branched processes. All physical processes (at the appropriate nomological levels) do still require space-time instantiations but do not require the continuity of the space-time substrate. Thus, if it is upheld that persons are metaphysically processes (similar to games or domino cascades), then many of the criticisms related to the MUP dissipate. For example, processes may occur across alternative substrates and may be replicated and, therefore, persons may ‘occur’ across alternative substrates and may be replicated. However, if persons are not processes (or only in part are processes), but rather substrates, then the MUP becomes less feasible.

The persistence problem, as it relates to the MUP (substrates and processes), allows for the following novel presentation:

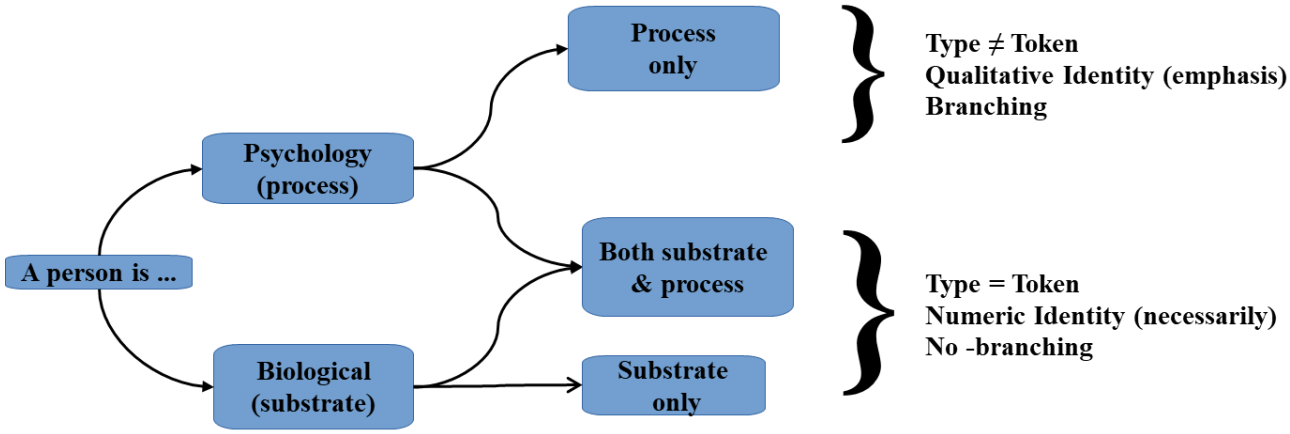

Figure 3-1: The persistence problem as it relates to the MUP

Three options (expressed from the bottom to top options within Figure 3-1) present themselves in relation to personal identity (PI) as:

(1) PI = Substrate. PI is constituted only by the substrate, which is the necessary and sufficient condition for continuity of identity (the biological solution).
(2) PI = Substrate + Processes. PI is constituted by both the substrate and the processes arising from it and both are needed for the continuity of the person (folk PI).
(3) PI = Processes. PI is constituted by processes arising from the substrate, which are the necessary and sufficient conditions for (partial) continuity of identity (the psychological solution).

Option 1) may be queried on the basis of the corpse problem (also presented in the previous chapter and to be more fully developed in the current chapter in section 3.5). In essence, the problem may be stated that if a person is no longer present when the biological body remains, then in some sense a person is not the biological substrate. Furthermore, those who wish to retain the view that persons are identifiable with the substrate only (e.g., Olson 2004; Blatti and Snowdon, 2016), must face the challenge of persons without minds, which is counter-intuitive to our under-

standing of persons. This dual problem of persons without processes (the corpse problem) and persons without minds is a theme throughout this chapter as the current thesis evaluates the substrate-only position.

Option 2) avoids some of the difficulties of option 1) by asserting that a person includes the processes (the mind) but *is* (identity claim), nevertheless, inextricably bound to the particular substrate. The thesis refers to this as folk personal identity (folk PI) as the common way that persons are categorised and identified within day-to-day life and this view will be discussed further in section 3.4. The need to bind the person to a particular substrate raises further problems, in that either this assertion is arbitrary (see previous discussed in section 3.2) or it requires that the substrate has unique properties or processes (to be evaluated further in section 3.6). In essence, to claim that a substrate has unique properties or processes is problematic since a unique physical property is contra-indicated by the empirical trajectory of sciences, where multiple realizability is more likely, and uniqueness of processes may be pragmatic due to complexity of replication, which implies that this is a pragmatic (engineering) problem, rather than a metaphysical one (which is the focus of the current thesis).

For both 1) and 2), if the person *is* (identity claim) the substrate (e.g., our biological body), then the MUP is not feasible as the replicated artifact, while instantiating the same processes would not be the same person and, therefore, the 'upload' would not have occurred. This same line of reasoning applies if the person *is* (identity claim) both the substrate and the processes of that particular substrate. In both of these options, the person is identified (either wholly or partly) with the space-time continuity of the substrate (the biological solution).

Option 3) asserts that the person is not identified with the substrate (fully or partially), but rather a person *is* (identity claim) the processes that the substrate instantiates. If these processes were to be replicated in another substrate, then the same person would be instantiated. This is akin to the psychological solution and equates the person to the mind. Precisely what these processes making up the mind are, is yet to be determined and competing theories have been discussed in the previous chapter. The precise nomological level (although important for the MUP) is not asserted in the current chapter, which simply notes that, in relation to persons, the psychological solution asserts that the mind (however constituted, at whatever nomological level, is needed for replication) is the necessary and sufficient condition for the continuity of persons. As will be demonstrated later (section 3.4), the psychological emphasis allows for instantiation of the psychological across substrates and, therefore, allows for the possibility of an artifact to instantiate the same person. The concept that a person is a process, rather than an entity, is further developed in sections 3.5 and 4.2 where the corpse problem acts as an argument in its favour.

The options presented in Figure 3-1 are associated with various philosophical concepts, which will be addressed throughout the current chapter. First the chapter focuses on the intuitions of whether persons are primarily bodies (substrates) or minds (processes) in terms of a discussion of famous thought experiments. Then, the chapter further develops relevant themes of numeric and qualitative identity (section 3.5.), type and token distinctions (section 3.6), as well as branching and non-branching distinctions (section 3.7). Each of these concepts is to be elaborated on within the relevant sections. The three options depicted in Figure 3-1 are presented here to orientate the reader as a lens for the MUP and will be referred to throughout the chapter.

3.4 Overview of thought experiments with elicited intuitions

Within this section, the thesis overviews some of the dominant thought experiments that have been introduced into the persistence of identity problem. The aim here is to explore some of the boundaries of intuitions in relation to personal identity as it relates to the MUP. Because the MUP allows for an artifact that has discontinuity of substrate (within space and time), and the biological solution's primary argument is the need for continuity of substrate in space and time (see e.g., Wiggins, 1967), this is the primary emphasis in this section's evaluation. Furthermore, a criticism of mind uploading within the philosophical literature is that a breach in space-time continuity of substrate results in the termination/cessation of persons (e.g., Corabi and Schneider, 2012; Cappuccio, 2017). In the previous chapter, the mind was identified with different nomological boundaries of processes (external and internal, higher and lower level processes). However, the current chapter is not concerned with the necessary and sufficient nomological boundaries needed for mind replication (e.g., whether the whole substrate or part of the substrate is needed for psychology to be instantiated), but rather to contrast the psychological (however achieved) with the biological. If a person is the same as their mind, then the psychological stance is upheld (irrespective of nomological boundary needed) and the MUP may be feasible. However, if the mind is insufficient for personal identity and the space-time continuity of substrate (the body) is necessary, then the MUP is contra-indicated. The distinction between substrate and process is, therefore, continuously emphasised within the current section with the biological associated with the substrate and the psychological with the processes.

As will be demonstrated in this section, much of the literature on the persistence problem relates to thought experiments that aim to elicit intuitions in support of the theory being posited. Although intuitions themselves are not sufficient to validate a metaphysical stance, they are presented here to indicate the nature of the problem as presented through the manifest image. The general structure of the section is to first present how personal identity is commonly identified (here

termed folk personal identity). From there, cases of body or mind swapping are considered, which serves as an overview of the primary distinction between biological and psychological solutions. Then the focus turns to whether a particular part of the biology is more important (i.e., the brain). The idea of perfect replication (molecule for molecule) is considered and then partial replication (e.g., replication in non-biological substrates). Finally, these are compared with possible MUP scenarios of gradual or scan-copy replication (see section 1.4 MUP matrix). As stated earlier, the emphasis is on what is necessary and sufficient for persons to continue and whether the same substrate (identified through space and time) is necessary for persistence of identity.

Human intuition about personal identity appears to be primarily orientated to the psychological view (Nichols & Bruno, 2010) as well as to be influenced by how the thought experiment is posed. For example, if a thought experiment were to begin with "you wake up inside the body of another person", the snuck premise is that "you" are not the body you inhabit (however defined) and, therefore, is likely to introduce a bias in a particular direction. Consider if this were applied to the MUP. If one were inclined to assert that the MUP is feasible, the scenario may begin with "you wake up and look around to see a robot reflection staring back at you the mirror". Alternatively, if one were inclined to assert that the MUP is not feasible, the scenario may begin "A replica with your memories wakes up ...". Therefore, a caution in how one words the scenario is important to ensure that a snuck premise is not present. This section's agenda is to present the thought experiments in as neutral a manner as possible to evaluate various thought experiments. The question of personal identity in these thought experiments does not contest feasibility of replicating minds/persons and, therefore, that a person (of some sorts) is instantiated is assumed here. The question is rather whether this person who awakes is in fact 'you'.

To begin with, it may be queried as to what the manifest image of a person is at present and how personal identity decisions are made. Setting aside the possibility of solipsism (that a single person exists surrounded by non-persons), let it be assumed that the world is populated by persons; that each of those we meet and interact with are in fact persons. Currently we identify persons primarily based on their appearance and behaviour. Both visual appearance and behaviour alter over time, yet we attribute the personal identity based on a continuity of these conditions. When a friend enters the room, prior to any distinguishable behaviour, we identify the biological entity as our particular friend. Now consider that you have heard that the friend has been in a horrific accident and their physical conditions are such that they are now unrecognisable. You meet this person and as the person sits down they begin to talk to you and you recognise, based on the verbal behaviour (e.g., "do you remember when we ...") that this is indeed your friend despite the

discontinuity of appearance. What if your friend were an identical twin, how would you have identified your friend form their twin? Each twin may have a different personality (way of interacting with their environment) as well as having distinct memories, and so it is probable that you would rely on these psychological differences to tell them apart. Consider the following premises:

(a) Continuity of biology (visually identified).
(b) Continuity of psychology (behaviourally identified).

Each of these premises may be further developed. For example, within a), a greater biological distinction can be made, in that even though the observable appearance may be the same (such as in twins), detailed biology would be vastly different (e.g., variation in epigenetics). This variation in biology would impact on variation in psychology (by the supervenience principle). Therefore, the initial appearance of everyday identification may be superficial and 'biological' may be further expanded to include all observable physical conditions of the substrate. Within b), a greater psychological distinction may be asserted, as internal processes (not just the externally observed behaviours) may also be included into the psychological category. This could include, for example, memories that we do not have awareness of (e.g., non-conscious memories) that nevertheless impact on our cognitions (e.g., conscious thoughts). In this chapter, the stance that both biology and psychology are needed is termed folk personal identity (folk PI) and is aligned with the second integrated option (as put forward in Figure 3-1 formulated as PI = Substrate + Processes. For folk PI there is a necessary and sufficient continuity of both biology and psychology.

The thesis now turns to the idea of body/mind swapping, which calls into question the folk PI intuition PI = Substrate + Processes. The idea of body swapping was grounded in philosophy by John Locke (1689), although religious texts have many instances of bodily control being attributed to persons other than the traditional 'occupant' of the body (e.g., In Mark 5: 1-20 of the New Testament, a particular man's body is said to have been inhabited and controlled by a multitude of evil spirits). Within Locke's work, he presents the idea of a prince waking up in a cobbler's body and *vice versa*. In the first scenario, the cobbler's body embodies, in some sense, the psychology of the prince (which for Locke is focused on consciousness that, in turn, is primarily based on memories). Other persons see the body of the cobbler but within the mind of that body are the Prince's memories. The thought experiment relies on the intuition that a person is the mind (that which retains the memories) rather than the body that is instantiating these memories. The intuition that is asserted if one accepts the thought experiment is that bodies (space-time continuity of substrate) are not essential for personal identity.

Williams (1970, 1973), who emphasises biological continuity, counters this thought experiment with one of his own where he believes the essential conditions of the body swap experiment remain, but the interpretation (intuition) can be changed. In his thought experiment, instead of a person waking up in another body (which may include a snuck premise of psychology as persons), we are to imagine a person through some procedure being altered into thinking that they are another person. The psychology of the person prior to the procedure is completely different from the psychology of the person after the procedure. Williams aims to create a similar context of the traditional body swap thought experiment that posits two persons (person A and person B) who are both manipulated into thinking that they have the psychology of the other. Who is which person now? In this scenario, a person can be thought of as being driven insane to believe that they are another person, despite remaining as the person identified with the original body. To up the ante, Williams proposes that the participants are told that the body at a later stage will be tortured and that this threat of torture is likely to illicit more accurate intuitions (i.e., for Williams that the person prior to the swap will not want the body that they currently instantiate to be tortured irrespective of altered psychology). This leads Williams to assert that identifying the correct person would be a matter of following the body and that the personal identity of person A would remain with body A despite the psychology now being different (that of person B) and *vice versa*. If this is the case, then according to Williams, variation in psychology is not a sufficient criterion for determining personal identity, whereas biological continuity is.

The presentation of both these thought experiments indicate how easily an intuition can be snuck into the premise. In both Locke's (1689) "should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler " (pp. 250-251) and in Williams (1973) "I shall undergo significant psychological changes" (p. 53") and "I shall not remember ..." (p. 52), the wording of the thought experiments implies the conclusion. Here the thesis attempts neutral language and the basic argument may be presented as follows: Premises 1 and 2:

(1) Body A and Body B instantiate Psychology A and Psychology B respectively.
(2) Due to some procedure Body A instantiates Psychology B and Body B instantiates Psychology A.

Depending on the initial intuition, the following conclusions can be reached:

(a) If persons are attributed to the psychology, despite substrate (body) variation, then persons are necessarily and sufficiently psychological phenomena (psychological solution).
(b) If persons are attributed to the substrate (body), despite psychological variation, then persons are necessarily and sufficiently biological phenomena (biological solution).

It is unclear as to whether there is a clear way to resolve these contradictory intuitions from within the logic of each thought experiment. However, the above thought experiments establish the primary difference between the biological and psychological solution to the persistence problem, namely, if one wants to establish what matters for the continuity of persons, one can either follow the body or follow the psychology.

Precisely which aspect/s of psychology should be emphasised as necessary and sufficient condition/s for continuity of persons is beyond the scope of the current thesis and there are many options to pursue (e.g., personality, desires, beliefs, and so on). In Locke's (1689) perspective, memory was the primary condition, which has elicited much debate. Reid (1785/1941) noted that memory is a poor criterion as memories fade and change over time. The thought experiment of the brave officer (ibid.) was put forward where we are to consider a person who performed actions as a child (robbed an orchard for which he was flogged), and as he grew up, this child became a brave soldier who received commendations for bravery (taking the standard). At this time, the soldier remembers the actions of the child. The soldier grows older into an old man and becomes a general who can no longer remember the child's actions but does recall the actions of the brave soldier that he once was. If memory is the criterion for determining personal identity, it would have to be stated that the old man is a different person than the child (as there is discontinuity of memories).

One way to respond (e.g., Parfit, 1984), is to state that what is necessary is not that memories are retained, but that there is a continuity of memories. The young soldier remembered the child's actions and the old man remembered the brave soldier's actions. What is important for the memory condition is not that all memories are retained (most persons cannot remember most of their lives), but that they were at some point accessible to the person and that there is continuity (not accessibility).

Another criticism of the memory criterion is that it appears to present a circular argument and may be presented as follows: For person A to be the same as person B, and for person B to be the same person as A, requires remembering person's A experiences, however the act of remembering (that a person can remember their experience) assumes they are the same person. Therefore,

it is argued, that the memory condition is circular, in that the act of remembering assumes the same person and the criterion of remembering asserts the same person. In essence, can a person remember an event that did not occur to them and, if so, would this not be a false memory?

In response to this circularity, Shoemaker (1959; 1970/2008) develops the notion of quasi memories (Q-memories). Q-memories do not require the assumption of the same entity (physical substrate) experiencing them, but rather defines memory as a causal link between the memory and the past event. Q-memories then may occur across substrates. The memory condition then can be adapted as follows: person B is the same as person A, if they (person A) have sufficient Q-memories of person B. The Q-memory, thus, focuses on the causal link between memories (this could, in principle, be expanded to all psychological attributes), rather than the physical entity that was present at the time. Put within the terminology of the MUP, it is not the substrate that matters but what processes are instantiated by the substrate. One reason why this causal continuity is preferred over purely physical consistency, is that this is already the case within our current biological processes. The memories of one's childhood occurred to different physical matter (our atoms and cells change over time). Therefore, in relation to memory, what matters is not the substrate but the continuation of causal relations (processes). This view is congruent with neuroscience (e.g., Kandel, 2007), where memories are active neuronal patterns (i.e., they are processes performed by neurons) and, therefore, should the process be performed by an alternative system the same memory (Q-memory) would be instantiated.

The thesis now considers thought experiments that focus on the brain rather than the whole substrate (the brain criterion). Shoemaker (1963, 1984) put forward a thought experiment where the brains of two persons are swapped into each body respectively. Mr Brown and Mr Robinson undergo brain surgery where Mr Brown's brain is placed in Mr Robinson's body. The new creature can be called Brownson. Furthermore, let it be assumed that Brownson has all of Mr Brown's mental life (the neurocentric assumption). Now, which is the donor and which is the recipient? Who is Mr Brown, or is there no more Mr Brown and only the new person Brownson? If one holds to the view that the body is the person, it would be said that Mr Robinson received a new donated brain. If it is held that the mind is the correct condition for personal identity, then Mr Brown is the brain (particular part of the substrate) and recipient of a donated body. The general intuition for Shoemaker is that the person will be attributed to the brain rather than the body and this experiment is, therefore, an application of the neurocentric view of the mind and persons (persons are instantiated in brains).

This thought experiment does not directly contradict the biological solution (although it may contradict embodied cognition), because the biological brain of each person is retained (space-

time continuity of substrate part) as a necessary and sufficient condition. However, a psychological condition may be implied when considering why it is that the brain is important for the retention of personal identity. If brain retention is necessary for persons, it is likely that this is because of the processes that the brain serves (otherwise loss or replications of other organs, such as a heart, would also imply loss of person continuity). Therefore, brain retention (as opposed to any other substrate part) as the necessary and sufficient condition implies that what the brain does is essential to continuity of persons. As the processes of the brain are associated with the psychological, it can be inferred that personal continuity is maintained through psychological continuity. Furthermore, a brain may be retained in a corpse and yet the person does not remain, which further indicates that it is the brain's processes that matter (see section 3.5 on the corpse problem).

Consider that brain processes may be currently stopped and restarted (breaching temporal continuity), termed 'brain freeze' in the current thesis. For example, Hayworth (2010) discussed temporal discontinuity such as when a medical procedure (e.g. deep hypothermic circulatory arrests; see Reich *et al*., 1999 and McCullough *et al*., 1999) may require the cessation of neuronal processes to administer surgery. In these instances, there is no query in relation to the cessation of the person once the biological processes are restarted. Brain freeze, therefore, states that temporal discontinuity of brain processes does not negate personal continuity. If this is the case, then the proponent of continuity of the brain substrate part as essential is only left with the condition of space continuity as the essential condition for personal identity. The current thesis now turns to a further complexity, the cases of fission and fussion, which emphasises spatial discontinuity.

Fission cases (Lewis, 1987, Parfit, 1971, 1984) consider what the consequence to personal identity would be if a person were to be split in two (or more). There are multiple ways for fission to be presented (e.g., Parfit's teleporter later in this section) but, for the present, consider the two halves of a brain that may be retained separately. For example, consider a person who is in an accident that results in the loss of the left half of their brain. It appears natural in this instance to assume that the person will persist despite the loss of the left half of their brain. Now consider what would have happened if that same person were to have lost the right half of their brain. Again, it appears that the person will persist (now retaining the left hemisphere). What would happen then if the left half (Lefty) and the right half (Righty) were separated and placed in different bodies? If the neurocentric view is assumed, as in the brain swap thought experiment, it could be said that the person has now separated into two of the same person (fission). The argument may be presented as follows:

(1) Person x at t_1 persists at t_2 if there is continuity of the left and right hemispheres of the brain (brain continuity).
(2) Person x continues at t_2 if retaining only the left (or right) hemispheres.
(3) If retention of either left or right hemispheres is sufficient for the persistence of person x, then, if both are retained separately, then person x persists separately into two spatial locations.

In 3), person x fissions into two substrates both with equal claim to be person x. Furthermore, if no particular space-time continuous biology (either Lefty or Righty may be used) can be non-arbitrarily assigned to the person x, then no particular space-time substrate can be an essential necessary or sufficient condition to the personal identity of person x.

Although ideas of gradual replacement are addressed later in this section in more detail, consider here the possibility of a functional equivalent artifact of Lefty being attached to the biological original substrate Righty (and *vice versa*). Then, at some point, the biological original Righty is destroyed in an accident and only the artifact Lefty remains. If it is inferred that person x continues throughout all these transitions (and why should it not), the continuity of a substrate in time and space appears an arbitrary condition. In some of these cases of brain fissioning, there are now two separate numeric whole substrates (e.g., if Lefty and Righty occupy different bodies) that have equal claim on the personal identity of person x. What appears to be the case, is that person x has split and branched into two person x's to be discussed later (see section 3.7) and, therefore, that numeric identity and the claim of uniqueness in relation to a particular substrate may be queried (further developed in section 3.5).

This line of thought has been further extended (Nagel 1971/2008) in discussions where current brain bisection surgery (where the corpus callosum is severed to ease epileptic symptoms) indicates that the two hemispheres operate independently to some extent. An example of Nagel (ibid.) is that of a pipe being placed in the right hand of such a patient and then the patient being asked to write with the left hand what the right was holding. The patient begins slowly with the letters P and I and then proceeds to quickly write PENCIL. The one hemisphere is guessing (see section 2.5.2.2 in relation to confabulation), but then the writing hand deletes the ENCIL and proceeds to draw a picture of a pipe. Are there two (or more) people inside of each of us even now? For Nagel (ibid.), this led to the conclusion that the idea of a unified self is an illusion and that the "... simple idea of a single person will come to seem quaint someday" (ibid., p. 243).

If a person is an entity, a particular numeric thing, then a problem arises, in that there appears to be two persons within the same substrate currently. If each conflicting process is a person, then the mind consists of multiple persons. However, if persons are viewed as a unified process, then the difficulty dissolves. At any given time, a unified process may have interchangeable entities (e.g., use different neurons), have multiple redundancy processes (e.g., neural plasticity), as well as have relatively autonomous sub-processes or modules (for example see Dennett 1991, 2017; Minsky 1988, 2006). If persons are unified processes, the idea that parts of the brain (and their functions) work independently when separated and cohesively when connected is simply the nature of neural functions and offers no such conundrum as the person never was a thing (a numeric entity) but rather that the person is the unified collective processing of multiple functions.

Fussion cases are similar to fission cases, but in fussion Lefty from person x and Righty from person y are conjoined in one body (either the person x, person y, or some other substrate). Are there now two persons in one body, or does the fussion interact to form a new person xy? Furthermore, how would these identification options be made? Consider what would be the relation to personal identity if *only* one psychology (e.g., the psychology previously of Lefty) were to be present? If one psychology were to be retained (despite a further hemisphere being active), it would seem that this would be the personal identity that is retained. If psychology from both were to be present (e.g., memories from both Lefty and Righty), then it could be argued that two persons survive. If, however, a novel psychology were to emerge (e.g., a different personality with different memories from both Lefty and Righty), then a new person would be said to emerge (call this combination Lighty). It seems that what both fission and fussion indicate, is that the allocation of personal identity to substrates (here brains, or brain hemispheres) that occupy space and time appears arbitrary, and that it is rather psychological continuity that matters for establishing personal identity.

The idea of fission has been further developed in thought experiments that allow for a molecule for molecule replication of persons, such as in Parfit's (1984) teleportation thought experiment. Consider some time in the future that a person lives on earth but works on Mars. Every morning he steps into a teleportation machine that scans him, deconstructs the molecules, beams the information from the scan to Mars, and a perfect molecule for molecule replica is made on Mars where the person goes about their work, to be teleported back to earth (in the same manner) at the end of the work day. This type of teleportation is, therefore, similar to a destructive upload scenario. For years, the person continues in this way and then one day they step into the teleporter, push the button, and nothing appears to happen. The person steps out of the teleporter to tell the technician that the teleporter must be malfunctioning. He is then told that there has been a change in policy at the teleportation company and that the machine is working perfectly. However, instead of him being

deconstructed on this end of the device, he has simply been replicated at the terminal location while remaining intact at the departure location. This second type of teleportation is, therefore, similar to a non-destructive upload scenario.

The molecule for molecule replicant removes objections regarding what transitive properties between the variations in substrates would be required, i.e., from biological to synthetic substrates. The molecule for molecule replicant allows for an exact qualitative identity (there is no distinguishable quality in any way), while allowing for a distinction of numeric identity. In essence, the only qualitative difference (the only different condition) between the original and the replicant, is that substrates occupy different space-time locations. If there is no qualitative difference and only a numeric difference between the replicant and the original, the question of why the numeric matters is raised (see section 3.5). For Parfit (ibid.) (for now, as the thesis will return to his view in section 3.5, 3.6, and 3.7), what matters is that there is psychological continuity, and as long as this is maintained, then so is the personal identity (irrespective of the substrate that instantiates the psychology).

Davidson (2013) offers a related molecule for molecule thought experiment with different conclusions. Consider Davidson going for a walk near a swamp, when he is struck by lightning and dies. At the same time another bolt of lightning hits the swamp and, due to some improbable event, this interaction results in a creature emerging from the swamp that is molecule for molecule Davidson's replica (the swamp man). This replica acts the same and is indistinguishable from the original Davidson. Swamp man itself experiences no distinction between the psychology instantiated in the original substrate and the psychology now instantiated in the swamp man substrate. However, the swamp man cannot "recognise" Davidson's friends, as the substrate has not previously instantiated the initial cognition (although he/it remembers the events as if they were his/its own). The swamp man does not have the same causal history because its history only began at the time of the lightning bolt. The implication for the MUP, if the swamp man argument is upheld, is that where there is no space-time continuity of the substrate, there is no continuity of the mind/person (although an indistinguishable replication exists). That the swamp man has a different space-time continuity and history is not denied (indeed that is the parameter of the thought experiment), but the question remains as to why this matters. Just as one can assert that space-time continuity of substrate matters, another could assert that it does not.

Consider a functionalist assertion (that what matters are the functions) then; since the swamp man performs the same functions (as per the thought experiment), it *is* Davidson. True, there is no 'recognition' of friends but there is the same function of recognition (Q-memories) of friends and, in the functionalist assumption, this is what matters. Furthermore, if it is argued that

in order for the same psychology to result, there needs to be the same physical causal history (i.e., only the same continuity of substrate can result in the same continuity of psychology), then the thought experiment contradicts itself, as the parameters of the thought experiment are that the swamp man is psychologically indistinguishable from the original Davidson. So, either the swamp man thought experiment cannot metaphysically happen (i.e., there needs to be causal history of substrate for the same psychology), or it begs the question as to what matters.

So far, the emphasis within the thought experiments have been on alternative biological substrates, whether minds in alternative bodies, parts of bodies being moved about, or molecule for molecule biological replicas. Now the thesis explores some thought experiments of uploading with synthetic artifacts (the A of the ABC options). Dennett (1978b; 1982) introduces a thought experiment that integrates aspects of many of the previous thought experiments, while assuming a neurocentric and computational theory of mind. Consider that Dennett is asked to perform a dangerous mission (disarm a bomb), but the proximity to the bomb would destroy his brain. To resolve this problem, Dennett's brain (the substrate part) is separated spatially from his body (the substrate *sans* the brain) but connected though some form of information communication so as to control the body. When the body stands next to the brain stored in a vat and looks at it, where is Dennett? The question of 'where', implies a spatial question. Dennett (personal identity) can be seen as either the object standing (the body), the object in the vat (the brain), as being both objects but now occupying two locations simultaneously, or as the unified process that can occur either in the brain or between the brain and the body. Then Dennett is sent off to disarm the bomb, leaving his brain back in the safety of the base, but his body begins to malfunction and senses are lost (hearing, touch, and so on) until there is no connection between the brain and the body. It dawns on Dennett that he (personal identity) is no longer an extended interaction between body and brain but rather now a bodyless brain back at base. Luckily they give him a new body and Dennett is back to interacting in his environment while his brain stays separate from this new body.

Up to this point in the thought experiment, the experiment can be seen to be a variation of the brain swap thought experiments, extending the question as to what it would mean if the brain and body were stored separately. The intuitions that may develop are that the self is essentially a control system (the brain controls the body and when the body ceases to exist the self is still retained) and that the location of the control system does not matter, but the interaction with the environment (what function the body performs) does.

Then the thought experiment takes another twist and it is found out that the brain was functionally replicated in a computer. A switch that connects the biological brain or the computa-

tional brain to the body was put in place and it is found that Dennett cannot distinguish any condition of the self, irrespective of which system is activated. The further intuition now is that the mind is a functional computational system (if the person cannot tell the difference between biological and synthetic substrates performing the same functions), as well as that continuity of functions across substrates would lead to continuity of self. This thought experiment aligns with the scan-copy upload (the brains functions are initially scanned and copied to produce the computer) and with the non-destructive uploads (both brain and computer continue to exist). The thought experiment ends with the two systems branching out of synchronicity and claiming that switching between the brain and the computer is switching between two separate Dennett persons who occupy the one Dennett body. This intuition leads to the branching view of persons, that if systems with the same functional history diverge at a later point, then two separate persons will emerge (further discussed in section 3.7). The self (personal identity), in this thought experiment is, therefore, a matter of continuity of psychology rather than continuity of any substrate as well as the psychology being constituted by functions and allowing for the possibility of branching.

The second thought experiment that allows for synthetic substrates is by David Chalmers (2014), where he transitions from BioDave to DigiDave through the gradual replacement of parts of his brain with synthetic functional equivalents. The original substrate can be called BioDave and relates to *n.* If one per cent of Dave is replaced, this would be Dave (n +1) until BioDave has been 100 percent replaced by DigiDave as Dave (100). The argument in relation to identity may be presented as follows:

(1) For all n < Dave (n+1) is identical to Dave (n).
(2) If Dave (n+1) is identical to Dave (n), then Dave (100) is identical Dave.
(3) Therefore, Dave (100) is identical to Dave.

If it is accepted that there is continuity of the person with continuity of functions when there is replacement of biological parts (e.g., the gradual replacement of cells that bodies currently undergo), why should the continuity of person not occur if these parts are artificially replaced within a synthetic substrate? The assumption is that if a computer of sorts (DigiDave) can perform the same functions (functional invariant), then the same personal identity will continue.

The reader is reminded of the basic scenarios of the MUP matrix (Figure 1-1, section 1.4.), which include destructive/non-destructive and gradual/scan-copy dimensions. First consider the gradual/scan-copy dimension. Both proponents (e.g., Chalmers, 2014) and critics (e.g., Corabi and Schneider, 2012; Pigliucci, 2014) of the MUP assert that there is little metaphysical distinction

between gradual and scan-copy, as at the end of both processes there is a technological artifact, distinct in space from the original, which is said to instantiate the original person. Compared with time, there is a possible difference between the gradual upload (as exemplified in DigiDave) and the scan-copy upload because in the gradual upload (e.g., the transition from BioDave to DigiDave) there may be no temporal breach, whereas in the scan-copy there may be a delay between the scan and the copy being activated. However, once the gradual upload is complete (e.g., there is only DigiDave), there is the possibility of scan-copy and temporal delays (e.g., replicating or turning DigiDave off and on at a later stage). So, although within the gradual upload of DigiDave no breach in time may occur within the scenario, the end result of DigiDave allows for this possibility. For Chalmers (2014), what metaphysically holds true for the gradual upload will hold true for the scan-copy upload.

Second, consider the destructive/non-destructive dimension of the MUP matrix (Figure 1-1). If, in the gradual and scan-copy scenarios, the end result is an artifact that could, in principle, be replicated multiple times, this leads to questions of whether persons are numeric, non-branching, unique types (see sections 3.5, 3.6, and 3.7). The idea that personal continuity may allow for multiple replicas is contentious and, to avoid this, the MUP may opt for destructive uploads (where only one space-time instantiation of a person is activated at any given time). The current thesis has viewed the idea of personal continuity being dependent on destructive/non-destructive options as reflecting arbitrary/extrinsic conditions that offer no challenge to the MUP (see section 3.2) and has preferred to emphasise intrinsic conditions. However, should the need for one space-time instantiation be upheld (which, nevertheless, does allow for breaches in space-time continuity of substrate), the MUP may simply opt for a destructive upload. Space-time continuity of substrate is, therefore, denied by the MUP as a necessary and sufficient condition for how space-time instantiations are affirmed.

The thesis now presents each of the thought experiments discussed above in Table 3-1, focusing attention on the distinction between the biological and psychological criteria. The psychological condition is simplified in one category and no attempt has been made to clarify the particular aspects of psychology being emphasised by each thought experiment. While the previous chapter elaborated on some of the difficulties of what a mind is and whether it may be replicable, here it is assumed that the mind (however defined) is replicable. Within the biological solution, even if a body is replicated (e.g., the molecule for molecule thought experiments), and even if this body replication results in a qualitatively identical mind, the person is denied as being replicated as, in this view, only the original body (space-time continuity of substrate) is the person.

	Biological (Substrate)		**Psychological (Process)**
	Space	**Time**	**(External & Internal; Hl & Ll)**
Folk personal identity	✓	✓	✓
Thought experiments			
Prince and the Pauper (body swap)	X	✓/X	✓
William's neurosurgeon (mind swap)	✓	✓	X
Brain transfer (brain swap)	X (partial)	✓	✓
Brain freeze	✓	X	✓
Fission and fussion (lefty and righty)	X (partial)	✓	✓
Teleportation (molecule for molecule)	X	✓/X	✓
Swamp man	✓	✓	✓
Computer continuity (Where am I?)	X	X	✓
Computer continuity (DigiDave)	X	X	✓
The MUP scenarios as thought experiments			
Gradual upload	X	✓	✓
Scan and copy	X	✓/X	✓

Table 3-1: Necessary and sufficient conditions for personal identity

That persons are intimately connected with minds may be evident (although not assumed), in that the majority of thought experiments assume the mind as a condition for persons. Thought experiments that intuit the biological solution (e.g., William's neurosurgeon and Davidson's swamp man) still assume, nevertheless, that persons may include minds, but argue that the particular person is preferably understood as being identified with a particular body (space-time continuity of substrate).

The emphasis within the current section has been on whether space-time continuity of the substrate is an appropriate necessary and sufficient condition for personal continuity, with time and space allocated further columns. In Table 3-1, ticks indicate that the thought experiment (the table rows) intuits that space/time/psychology (the table columns) are necessary and sufficient conditions, whereas crosses indicate that space/time/psychology are not necessary and sufficient conditions. If space-time continuity of substrate may be queried or called into question, it weakens the intuitive argument for the biological solution and, by implication, increases the likely feasibility of the MUP. Although there are multiple avenues to pursue to further evaluate the disparate intuitions, the current thesis emphasises in what follows, the philosophical themes of numeric-qualitative identity distinction, the type–token distinction, and the branching possibility, as they relate to the three options of PI in relation to the MUP (PI = substrate, PI = substrate +processes, and PI = processes).

3.5 Numeric versus qualitative identity (what matters)

As discussed in chapter 1 (section 1.5.4), identity may be broadly categorised as either numeric or qualitative. For the purposes of the current thesis, alternative forms of identity, such as relative or contingent identity (e.g., Geach 1973; Griffin, 1974; Gibbard, 1975; Evans, 1978; Tye, 2000), are subsumed under qualitative identity since both qualitative identity and these alternative forms of identity assert that identity may be retained if only certain conditions of the system (however defined) are maintained. Qualitative identity is, therefore, the acknowledgement that identity does not require a complete 1:1 correlation of all conditions and is akin to the notion of partial identity. The current thesis presents absolute identity (contrasted with partial identity) as a form of strict numeric identity, where identity is retained if *all* properties and processes, at *all* nomological levels, is maintained over time. Absolute identity is related to absolute continuity, previously mentioned in section 3.2. Although there is significant overlap (both relate to the persistence problem of persons), the distinction here is that identity (absolute or partial) relates more closely to the attribution of sameness and distinction (how is this phenomenon the same and how is it distinct from other phenomena), whereas continuity (absolute or partial) relates to the collection of properties or processes that would need to be maintained for such an identity claim to be made. In this sense, continuity is the collection of necessary and sufficient conditions under which an identity claim may be made.

Absolute identity is deemed to be incongruent with the persistence of personal identity through the following argument:

(1) For P_1 to be absolutely identical with P_2, all conditions ($P_1 = P_2$) need to be retained across time from t_1 to t_2.
(2) P_1 is different from P2 (P1 being defined as the person at t1 and P2 the person at t2) because, at the very least, the condition of being at t1 or at t2 is a different temporal condition.
(3) Therefore, P1 ≠ P2.

Premise one is the definition of absolute identity that asserts a strict numeric identity as described in Leibniz's law[36]. The second premise is the assertion that, at the very least, the condition of occupying a different temporal condition (t_1 and t_2) means that not all conditions are the same. Furthermore, if persons are dynamic systems (which claim needs to be retained for there to be an actual problem in the persistence problem), then, by nature, persons are constantly changing.[37] If the person has changed in any way, the person cannot be said to be absolutely identical and, therefore, the conclusion may be asserted that $P_1 \neq P_2$ (in the absolute identity sense).

The alternative to absolute identity presented in the current thesis is partial identity, which may be formulated as follows.

(1) For P1 to be partially identical with P2, some conditions need to be maintained across time, i.e. from t1 to t2.
(2) P1 is different from P2 (P1 being defined as the person at t1 and P2 the person at t2) yet retains some necessary and sufficient conditions for PI.
(3) Therefore, P1 = P2.

Within partial identity there is only a need for some conditions to be the same and, therefore, this is closer to qualitative identity. However, partial identity is broad enough to include the traditional notion of numeric identity (e.g., the numeric body that changes over time).

Presented here is a brief overview of some of the responses in the literature to the notion of numeric identity in relation to personal identity. The notion of numeric identity may be either accepted or rejected. If it is accepted, it may apply to the numeric body or the numeric mind (at

36 The indiscernibility of identicals (sometimes termed Leibniz's law, introduced in section 1.5.4), states that if x and y are identical, then every/any property that belongs to x also belongs to y (Forrest, 2020).

37 If one considers that at certain nomological levels (e.g. quantum fluctuations), entities pop in and out of existence, absolute numeric identity becomes less and less feasible.

issue in both cases is still partial identity, in that certain conditions are affirmed as essential, while others are denied). If it is the numeric body that matters, then the substrate is necessary for the continuity of persons and the MUP is not feasible (e.g., Olson, 2007). If it is the numeric mind that matters, so long as there is only one mind instantiation (the notion of numeric identity does not allow one to many instantiations), the MUP is feasible (e.g., Lewis, 1987). In these MUP scenarios, the continuity of persons is dependent on no further instantiations (non-branching) being made and relates to a destructive sequential (only one replicant at a time) upload of the mind. If numeric identity is rejected (resulting in accepting qualitative identity) as relevant, it may be on the grounds of accepting the philosophical concept of numeric identity but rejecting the relevance to personal continuity (Parfit, 1984), or on the grounds of negating numeric identity as a meaningful philosophical category (e.g., Geach 1973; Griffin, 1974; Gibbard, 1975; Evans, 1978; Tye, 2000). For those who accept the qualitative identity option, there is the possibility of multiple replicas retaining the same personal identity (branching).

First, the section explores the option of asserting that numeric identity is philosophically valid but that it does not matter to the persistence problem (Parfit, 1971, 1984). Parfit accepts the formal character of 1:1 correlation of numeric identity between the subjects at different times, and states that it does not admit degree.[38] Mental continuity (termed the R-relation by Parfit) is deemed to be the essential condition of persons (persisting through time) and Parfit states that mental conditions are asserted as not having this formal character (mental conditions may have degrees of one-many and many-one relations negating the one-one relation of numeric identity). Therefore, for Parfit, what matters to persistence of persons is not numeric identity, but R-Relations (i.e., the qualitative identity of psychology that matters).

An alternative psychological solution was put forward by Lewis (1987), who maintained that a formal 1:1 correlation did matter (numeric identity) but was not to be the primary emphasis in personal continuity. The argument is loosely summarised here as:

(1) P1 = P2 when both are at t1 and share all properties (numeric identity).

38 This relates to Parfit's (1984; 2016) view that persons may either be simple or complex (see also Scheffler, 1982). In the simple view, persons are not a matter of degree and persons are some 'further fact' beyond their psychological continuity. In the simple view, persons are, therefore, a unique 'thing' that is not reducible or divisible and, therefore, also relates to substance dualists' assertion of the unified mind/person, where the person/mind is a numeric thing. In contrast, the complex view admits degree and sees the person as a continuity of diverse and dynamic processes. The current thesis holds to the complex view.

(2) I-Relations are the relations between persons at different times, i.e., the relation between P1 and P2 (the same person at t1 and t2 respectively).
(3) R-Relations are the psychological relations between persons at different person stages t1 and t2 (R-Relations of P1 and P2 at t1 and t2).
(4) If the I-relation = R-relation, then the person is the same person, based on psychological continuity.

Point (1) affirms numeric identity, in that a particular person can be numerically identified with a particular person at a particular time (t_1). However, as persons change over time, this is not sufficient for answering the question raised by the persistence problem. To answer that question, Lewis introduces the idea of an I-relation (2), which is distinct from numeric identity but maximally relates persons across various person stages. The I-relation will answer what it means for persons to persist, while not being the same as numeric identity. From here (3), the R-relations, like in Parfit's case, are put forward as relations of psychological continuity and connectedness between person-stages. If the R-relation and the I-relation are the same (4), then it can be said that continuity of the personal identity (I-relation) is the continuation of the psychology (R-relation).

Although the precise nature of these arguments are of general interest, the current section focuses on the use of the notion of numeric identity within the persistence problem. For Lewis (ibid.), the 1:1 correlation of numeric identity can only occur at t_1 and, therefore, for persistence (survival) to occur, a phenomenon needs to occur across time (t_1 , t_2, t_3, … t_n) and, therefore, a choice needs to be made as to what it is that needs to continue (what matters). There is, hence, a move from numeric identity to the I-relation, which can be seen as a move toward partial or qualitative identity, which implies that what is necessary for personal persistence is not numeric identity itself, but certain conditions being met. Evident in both of these psychological solutions is that absolute identity is denied and that even when numeric identity is maintained at t_1 (*á la* Lewis), it is not the primary focus for persistence of identity, which has to acknowledge that change is inherent to any dynamic system. Both Parfit (1984) and Lewis (1987) assert that it is psychological continuity that is essential for personal continuity. If it is the psychology (and not the particular body) that is essential for the continuity of the person, then so long as that psychology continues, then so will the person. For both Parfit and Lewis, the psychology is defined by relations (R-relations) and this is, therefore, aligned with the current thesis view that psychology is a process (PI = process), a way that things hang together.

The thesis now turns to the biological solution's use of the notion of numeric identity (e.g. Williams, 1973; Olson, 2007), which is defined here as the substrate-only option (PI = substrate)

of the biological solution. For the biological solution, the psychology may be non-necessary and the biological substrate necessary and it is, therefore, noted that the biological solution does not affirm absolute identity either. If personal identity may be retained without the continuity of psychology, then the solution is not an argument for absolute identity, as only the biological part is asserted to be necessary in terms of an assertion of numeric identity at the nomological level of the body (whole substrate). This condition, space-time continuity of the substrate, is deemed, in this view, to be the essential condition for continuity of persons. If this solution, where the person is identified with a particular substrate, were to be upheld, then the MUP would not be feasible.

As substrates (at the nomological level of bodies) are made of matter (here loosely defined as that which can be clearly defined as occupying a particular location in space-time), the thesis now turns to the question of whether matter (substrates) matters (is a necessary and sufficient condition for identity), by evaluating the thought experiment of the ship of Theseus (The Internet Classics Archive | Thesues by Plutrarch; Gallois, 2016). Does a numeric (emphasising the physical continuity of parts[39]) or a qualitative (emphasising the continuity of ship design) matter in relation to identity? In the thought experiment of the ship of Theseus (as well as other philosophical examples such as the grandfather's axe), a ship's parts are gradually replaced to such an extent that, at some time (t_2), not a single part of the original ship is used. Call this the replacement ship. Here the intuition is that the replacement ship retains the identity of the original ship, and it is still the ship of Theseus, despite no original part being present. This is similar to our current experience of personal identity, in that we retain personal continuity, despite having new biological matter over time[40]. Returning to the ship, suppose that, at a later time (t_3), all the discarded parts are then reassembled and another ship is constructed in the exact design of the first. Call this the reassembled ship. This ship has all the original pieces as well as the design of the original ship. Is this now the ship of Theseus? Alternatively, we could suppose that all these original parts were reassembled to have a completed different design, such as a house. Orientating the ship of Theseus within the current chapter's emphasis on substrates and processes may be presented as follows:

---

39 If a ship's parts can be gradually replaced to constitute a ship, then it stands to reason that these parts could be replaced instantaneously and the ship will still be constituted. In this sense, there is no difference in gradual or sudden replacement and this may be related to the gradual or scan-copy upload scenarios of the MUP.

40 Proponents of the biological solution acknowledge that parts of the substrate are replaced in persons. For example, Wiggins states that "John Doe the boy is the same human being as Sir John Doe the Lord Mayor, but not the same collection of cells as Sir John Doe" (1967, p. 9). Yet, despite the acceptance of the numeric parts changing, the numeric whole (the substrate of the biological body) is, nevertheless, asserted as a necessary and sufficient condition for personal continuity.

The four scenarios framed with substrate and process are:

(1) The original ship at t1 consists of certain material parts (substrate 1) and the ship's design (organisational processes).
(2) The replacement at t2 consists of different material parts (substrate 2) and the same ship's design (organisational processes).
(3) The reassembled ship at t3 consists of the original material parts (substrate 1) and the ship's design (organisational processes).
(4) The house of Theseus is a house (alternative organisational processes) with the same material parts as the original ship (substrate 1).

Solutions to the problem of identity can be presented as follows:

(a) Follow the material parts (the substrate parts).
(b) Follow the ship's design (the organisational processes)
    (b.i.) With numeric identity
    (b.i.i.) Without numeric identity.

The solutions presented here are intended to mirror the biological and psychological solutions as well as this thesis's emphasis on substrates and processes. Within a), there is a numeric identity of parts and the identity claim of the ship rests on retaining the material of the original ship. The problem with this view is that, within 2), the ship of Theseus disappears along with its parts. If this view is upheld in relation to persons, it would then be said that persons disappear along with their numeric parts (e.g., atoms, cells, molecules, and so on). Furthermore, in scenario 4), the house of Theseus would then be actually the ship of Theseus. Although numeric identity may be asserted in relation to parts (we could, in principle, locate and track the parts in space time), it can be said to offer significant problems for personal identity or identity across time.

Now, consider following the ship's design to establish identity. The design is understood here to be the instantiation of how things hang together and is, therefore, a process. If one upholds numeric identity, it can be asserted that there is one possible design instantiation (as based on the 1:1 correlation of numeric identity). In this instance, the ship of Theseus remains from 1) to 2), as there is only one instantiation of the ship design. However, as the reassembled ship is instantiated in scenario 3), a problem emerges as there can only one identity claim, yet 2) and 3) appear to have an equal claim (both have the same design). A closest continuer epistemology could be applied, but

there are now further complications as to which is deemed closer, as 2) has occupied a continuous design, whereas 3) is constituted by the original material parts. For 2), there is, therefore, a continuity of space-time location of the design, whereas in 3), there is a continuity of the space-time location of the parts (both may have a numeric claim for identity depending on where the preferred emphasis is placed).

It is, therefore, unclear as to which condition (numeric space time continuity of design or parts) is "closer". Furthermore, if 2) has an uncontested claim to identity (similar to how we have an uncontested claim to our personal identity yet replace molecules continuously) prior to scenario 3), and on creation of 3) becomes contested, then the nature of personal identity is arbitrary (dependent on extrinsic events). Therefore, the insistence on numeric identity creates problems for identity if one looks to the material that constitutes the phenomena (here the ship) and also if one considers the continuity of the organisation of these parts. However, if one holds to the ship design without holding to numeric identity, many of these problems can be solved. The ship of Theseus is both 2) and 3), as both instantiate the design. In identifying the ship of Theseus with the processes (here the organisational design), there is no problem of identity and it allows for the multiple realizability of ships (branching). The agenda here is not to resolve these long standing philosophical issues, but to show how assuming numeric identity leads to certain difficulties, as well as querying whether qualitative identity should be preferred.

Numeric identity, when applied to the biological solution to the persistence of persons, has two primary assumptions. First, is the assumption that there is only ever a 1:1 correlation of persons (as opposed to a one to many possibility). As stated by Williams: "the principle of my argument is, very roughly put, that identity is a one–one relation and no principle can be a criterion of identity of type T if it relies only on what is logically a one-many or many-one relation between things of type T" (1973, p. 21). By asserting that personal identity is a 1:1 relation, Williams aligns his views with numeric identity.

Second, the biological solution, once assuming a 1:1 relation, then asks what kind of numeric thing/entity/substance in relation to persons could account for this numeric identity? For the biological solution, the candidate is the substrate or as Wiggins stated, "... persons are substances" (1967, p 54). The material substance (the substrate) related to persons may be an intuitive assumption as the answer to what could account for the numeric identity, in that it is easy to identify a numeric thing with a numeric substance as wherever the singular person (in terms of the numeric condition) is in space time there is the substrate (the material substance). The primary condition for the biological solution is, therefore, that the whole substrate occupies continuity in space-time. This condition is unique, in that singular entities occupy unique continuity in space-time. For example,

think of any object such as a cup. The cup can be moved in space-time but retains space-time continuity (e.g., moving the cup from the kitchen to the lounge) and, therefore, there is a unique condition of space-time continuity of the cup entity and substance. The assumption of the biological is, therefore, that persons are things/entities (as opposed to how things hang together, processes). The dual assumption of the biological solution may be presented as:

(a) There needs to be numeric identity to personal continuity.
(b) Persons need to be things (substrates/substances).

In retaining these dual assumptions, the biological solution may be retained as a viable approach to the persistence problem, while denying either of these two assumptions implies the biological solution may be refuted. In essence, a person, in this view, is a thing or an entity of sorts and it is claimed that there is only one of this thing that can be numerically identified (using the space-time continuity of substrate as a means to identify the thing and distinguish from other things). Therefore, in this view, there can only ever (in all possible scenarios) be one of you and the metaphysical nature of you is a distinct numeric thing. If these two assumptions are retained, the feasibility of the MUP may be refuted but, if these two assumptions are denied, the MUP may be deemed feasible.

Presented here is a brief argument from numeric identity that may be presented from the biological solution to personal identity:

(1) Persons are entities.
(2) Bodies are entities.
(3) Persons and bodies share some space time property or process.
(4) Either persons ≠ bodies OR persons = bodies.
(5) Persons = bodies.

Both (1) and (2) assume that both persons and bodies are numeric entities (the numeric identity assumption) and relate to things that occupy space and time (space-time substrate continuity). This means, for this view, that persons are somehow substance sortals of some kind (Wiggins, 1967). Premise (3) is simply the observation that, throughout human history, where there is said to be a person (behaving, thinking, and so on), there is the observation that the body occupies the same space and time. In relation to (4), the proponent (e.g., Olson, 2007; Snowdon, 2014) of the biological solution may offer multiple reasons as to why the persons ≠ bodies option is rejected.

For example, if one accepts (1), (2), and (3), then we have two entities occupying the same space and time, doing the same thing, living a conjoined life of sorts, which sounds suspiciously the same as substance dualism. If substance dualism may be rejected (for reasons such as mental causation and over-determination discussed in the previous chapter) then, according to this view, the person must be the same as the body (5). Therefore, it logically flows from the assertion that both persons and bodies are entities that they are the same entity (or suffer the threat of substance dualism).

One version of the above type argument is Olson's (2007) thinking animal argument, (also named the too many thinkers argument) which has become the "standard argument for animalism" (Blatti, 2012, p. 685). The basic argument may be presented as follows:

(1) You have mental properties (thoughts).
(2) Your body (the animal) has mental properties (thoughts).
(3) "How could you and the animal have different thoughts?" (ibid, p. 29) (numeric identity restraint), therefore,
(4) You are (numeric identity) the animal (the body).

The two entities that are compared are the entity 'you' and the entity 'the animal', and it is determined that if both have the same thoughts, either 'you' are the animal, or there are too many thinking entities performing the same thoughts. Because, in this view, too many thinkers can be rejected based on Occams' razor as well as over-determination of entities performing the same processes (thoughts), it must be accepted that you are the animal (1:1 numeric identity between 'you' and 'the animal'). If one holds to the assumptions of numeric identity and two entities of 'you' and the 'animal', then the logic of the argument is powerful.

However, if one has alternative assumptions, then the argument becomes a category mistake between things and how things hang together, as well as the need for numeric identity. For example, assume that a psychological solution of persons is accepted. In this case the person, the 'you', is the psychology. Furthermore, if the psychology is defined as the processes of the substrate (and not the substrate) as emphasised in the previous chapter, the assumptions here may be stated that you = psychology = processes and, therefore, through transitivity, that you = processes. Now let us return to the argument of too many thinkers replacing the 'you' with processes and the animal with substrate:

(1) Processes (you) are processes (thoughts).
(2) The substrate (the animal) performs these processes.

(3) How could processes (you) and the substrate (the animal) have different processes?
(4) Therefore, the processes is the substrate.

The category mistake of asserting that 'you' are something more than the processes the substrate performs becomes evident. Premise 1) can be seen as stating simply that processes are processes and it is only in making 'you' something further than the processes that a problem begins to emerge. Premise 2) is another way of stating physicalism (of all kinds), where thought is the result of physical entities (however defined) interacting and, therefore, may be accepted by all physicalist stances and is not contested by the current thesis. If the substrate/process distinction is upheld, (3) becomes non-problematic, as the processes (you) are the processes of the substrate by definition. As there is no problem of (3), the leap to (4) makes little sense at all. Consider an analogous argument of a plane (entity) and flight (the process). The 'flying plane argument' may be stated as:

(1) Flight flies.
(2) Planes fly.
(3) How can both flight and planes fly?
(4) Therefore, planes are flight.

By retaining the substrate/process distinction it is shown how the logic of the argument can lead to an absurd conclusion.

Within the mind uploading literature, Corabi and Schneider (2012) follow a similar line of reasoning by asserting that the self/person is a metaphysical object (a numeric thing), identifying that object (entity) with the biological substrate and, therefore, denying the possibility of uploading. Furthermore, for Corabi and Schneider (ibid.), to upload requires a 'transfer' that has the notion of physically moving an entity (substrate space-time continuity) that is in contrast to the use of the term among standard computation, where 'transfer' is better understood as replication in alternative computers (substrates). Their argument may be presented as follows:

(1) The self is an entity.
(2) To upload requires the transfer of this entity.
(3) Replicating the mind is a duplication and not a transfer (based on numeric identity).
(4) Therefore, mind uploading is not feasible (at best it replicates but does not transfer).

The first premise is the assertion that the self is an entity (making the category mistake of conflating processes with entities). The second premise is a restating of their particular use of the term 'transfer', which relates to entities (rather than processes). The third premise asserts numeric identity as the condition for personal continuity. These premises lead to their conclusion that the MUP is not feasible. If any of the three premise are queried (as has been done in this chapter), the conclusion dissipates.

To confuse entities for processes may be a category mistake but are there reasons to assert that persons are processes (also see section 4.2). An argument in favour of persons being identified with processes (as opposed to substrates) may be seen in the corpse problem. Although others have identified the corpse problem (Shoemaker and Strawson, 1999; Baker 2000, 2005), the current thesis presents a variation of the corpse problem in relation to the MUP categories of substrates, processes, and persons.

(1) A corpse is not a person.
(2) A corpse is a substrate (body).
(3) Therefore, a substrate is not a person.

This is a negative argument against the person being the substrate[41] and may be seen as an argument against the body solution (animalism) that the person is the substrate. There are different responses to this problem (e.g., Olson, 2004; Sauchelli, 2016). For example, Sauchelli (2016) makes the distinction between persons as substance sortals as opposed to phase sortals. In arguing that persons are phase sortals, where the person phase is when the substrate is active and the corpse phase is when the substrate is inactive, the numeric identity of continuity of body is still retained. What is of interest, is that this reaffirms that persons are essentially (at least in part, in this view) processes. Although (1) may be denied by asserting that corpses are persons, it is unclear how a corpse is a person in any meaningful way (e.g., it is assumed that the corpse does not think or feel, as if it does, burial and cremation would surely be cruel, monstrous events). The common use of

41 An alternative positive argument may be stated as follows: 1) A corpse is not a person. 2) A corpse may be defined as the cessation of processes. Therefore, 3) The continuation of processes is a person. In both of these formulations premise (1) asserts that a person is not a corpse and then the difference between a persons or a corpse is asserted to be that persons have processes, whereas corpses do not. From there, the conclusion that persons are processes or corpses are substrates without processes is concluded.

language (although not to be relied on as infallible) leads to the implication that corpses are not persons. For example, it is common to state in the presence of a corpse that "Paul is no longer with us/Paul is departed/ Paul is no more" and so on. Furthermore, all our references of what it means to be a person are bound to the notion of interaction. If a person is in fact a corpse, then at what point do they cease to be? Is it at the point when the brain atrophies, the flesh is decomposed, or the bones decompose? Olson (2004) is aware of this difficulty and, in defending the biological solution, he contrasts living from dead things as "the stability of living thing, is dynamic ... only engaging in constant activity ... all this comes to an end when the organism dies" (ibid, p. 268). There is an interesting move by Olson to embrace the idea that persons are somehow processes, but Olson reverts back to arguing that 'you' must be the body as the *what* (note the emphasis on the need for an entity) that is performing these processes.

As far as the need for physical entities (substrates) to perform processes goes, there is no disagreement or contraindication with any of the physicalist/naturalist stances. That animals are needed for persons at present is not contested. Currently persons are only animals and animals of a particular kind (here meaning homo sapiens). Placed in the vernacular of the MUP, persons are processes and, currently, only a particular type of substrate (the animal) can instantiate these processes. The strength of the animal argument and the strength of the corpse problem coincide when the physicalist premise (that only physical things can think) is accepted but the substrate/process distinction is maintained. If both of these premises of a) physicalism and b) processes as persons are accepted, then it logically leads to the idea that persons are multiple realizable (the multiple realizable argument of the previous chapter). However, this conclusion of multiple realizable persons faces difficulties if one assumes that persons are numeric identities (that there can be only one person in any metaphysical sense). How can what is multiple realizable be constrained metaphysically (not based on arbitrary extrinsic features but intrinsic features) to one identity?

In this section, the thesis has queried the relevance of numeric identity to persons and as a viable philosophical stance and now turns to the alternative stance of qualitative identity. The agenda here is not to exhaustively explore all possible qualities that may relate to persons but to explore some that may account for numeric identity intuitions. Let these be termed numeric qualities.

First is the necessary quality of persons requiring space-time instantiations. It can be noted that, throughout history, there has been only one (numeric) observable instantiation in relation to persons. There have been no recorded accounts of person mitosis (where a person fissions into two) and although persons are not continuously monitored, it is assumed that a person can be tracked through space and time without breach. That there has only ever been one instantiation (correlated

to one body) need not imply that there will always be one unique instantiation (e.g., further developed in Dr Hatefrills thought experiment in section 3.6.). Space-time instantiations (as opposed to space-time continuity of substrates) is another way of asserting the physicalist position that serves as the basis of the current chapter. It acknowledges that any metaphysically real phenomena (whether this is a property or process, causal or non-causal, and so on) is physical. As the term 'physical' (only partially understood by our modern science) at the nomological level of our daily life, requires some form of space-time instantiation, this is deemed a necessary condition for persons[42]. Consider many of the examples of physical processes, such as baseball games, domino cascades, wave functions, societies such as the Vienna circle, software programs, and so on. In each of these phenomena, they can be said not to exist without space-time instantiations yet the space-time continuity of the substrate/s are not necessary. The current thesis, therefore, makes no assertion against space-time instantiations, but queries whether space-time continuity of the substrate is a necessary and sufficient condition for persons. It is put forward here that it is in conflating these two concepts that a further category mistake (as opposed to the category mistake of conflating processes for entities) is made, which has resulted in numeric identity being attributed to persons. If space-time instantiations are affirmed, while space-time continuity of substrate is denied as to what matters for personal continuity, then many of the intuitions that lead persons to the biological solution may be accounted for. In short, it is a necessary and sufficient condition that persons have a substrate but not that persons need a particular continuous substrate.

Second, it is acknowledged within the persistence problem that persons are dynamic (they change over time) and they are open systems (change as they interact with the environment). This is accepted by both biological solutions, where the body changes over time, and psychological solutions, where the psychology changes over time. If something changes over time (here meaning the time frames of daily life), it is difficult to see how two substrates may instantiate the same person in space-time (as each substrate will change in relation to space-time interactions). This difficulty is what may support the numeric identity intuition. The current thesis develops an argument to show how dynamic processes may relate to new persons as well as relate to the same past person over time in the section 3.7 on branching. What is noted here, is that the metaphysical question being raised, is whether it is ever possible (metaphysics relating to all possible worlds) for a person to occupy two locations simultaneously. If it is metaphysically possible for persons (part or

42 Space-time instantiations assert any form of instantiation that requires a physical basis. For example, virtual reality is still instantiated in space-time (i.e., there is some physical computer following physical processes that the virtual reality is instantiated with).

whole) as dynamic-open systems to occupy two space locations, then numeric identity is non-essential from a metaphysical perspective. However, if dynamic processes are accepted as a necessary qualitative condition, it can be inferred that these processes (instantiated in space-time) serve a pragmatic function (numeric quality).

Consider another type of dynamic-open system, that of a machine learning program, which needs to interact in an environment (virtual or real world) to perform its functions/processes. The condition of operating in time-space instantiations (how the machine learning program operates) does not necessitate that the program has one numeric substrate, nor does it necessitate that the program cannot be instantiated in alternative space-time locations (the program may be stopped and transferred to an alternative computer/s substrate). Assuming that a person is an dynamic-open system, what holds true for one system, such as the machine learning program, is likely to hold true for another similar system (transitivity). The need for dynamic interaction with an environment in space-time is, therefore, a necessary condition for persons, however the need for this interaction to only be instantiated in a particular continuous substrate is deemed non-necessary (e.g. the person, like the machine learning program, could transfer to an alternative substrate).

Third, the quality of being an indexical representational system (see Putnam, 1975; Hofstadter, 1979) is presented. Within the analytic tradition, there is an emphasis on speech convention (common language) and what this indicates about the world as it is (metaphysics). Common language may assume that a person is a numeric identity partly through the use of distinguishing pronouns of the singular 'I', or 'me'. However, the use of pronouns also allows for some variations that appear to contradict a singular entity. For example, it is common to say, 'I said to myself' or 'I told myself to ...'. Furthermore, there are psychological conditions (previously mentioned in section 2.4.1), such as dissociative identity disorder (DID), brain bi-section surgery, and so on, that indicate that a strict singular view of the self may be queried. If one holds to the view that the self is a numeric entity (substance), this leads to complications as to how to understand a self that is, at times, both singular and divisible. However, if the concept of a qualitative condition of an indexical self-referencing system (a process) is introduced, it may account for the self being, at times, both singular and divisible. This relates to the concept of the self as unifying processes and will be further developed in the current thesis's novel theory of the efferent self (section 4.3). The self in this indexical space-time occurrences, therefore, need not refer to any "thing" but rather is a collective term for multiple processes (similar to Hume's 'nation' (1777/1975), or Minsky's 'society', 1988).

The self as an indexical representational system, orientates representations (representations such as "I", "me") to itself (is self-referencing), as well as orientating this representation

within an environment (indexical space-time instantiations). This quality of being a self-indexical representational system may account for numeric intuitions, such as the self being singular and being orientated in space-time, while denying the numeric identity claims (a numeric quality). For example, a machine learning program (see space-time instantiation comments above) that is self-referencing and orientated to space-time location may satisfy a qualitative condition that can occur across multiple substrates. If the person is akin to a machine learning program, this would account for the singular (unified representation) space-time instantiation.

Fourth, is the qualitative condition of control. Indexical statements, such as 'Hitler invaded Poland' or 'I transferred money to Berlin', expand the notion of the self beyond the substrate's location. In these instances, the entity is performing actions that extend their area of control (see section 2.5.1.3 on the extended mind and the current thesis's theory of the efferent-self in section 4.3). Furthermore, consider tool extension (e.g., writing with a pen). In these instances, it would not make sense to question the agency of the person due to an intermediate object (e.g., the pen) not correlating to the numeric identity of the body. What this indicates, is that the speech convention of 'I' and 'me' need not always refer to the numeric body but, rather, may refer to the quality of control. Consider the empirical experiments where one person's thought processes control a body part of another through human-to-human technology interfacing (Gage, n.d.). For example, the neural signals from person A's brain (outputs) can be used to control the neural pathways (inputs) of person B's arm. In these instances, person B may make statements such as, 'I am not moving my body'. This type of thinking can also be attributed to issues of passivity in schizophrenic patients (although in this instance there is no second person controlling them but a part of their mind that they are not consciously aware of), or alien hand syndrome (Leiguarda, *et al*., 1993). Returning to the experiment of control through human-to-human technological interface, these phenomena may offer evidence that control is a primary quality of the self (wherever there is control, there is the self).

To summarise, a person needs to have qualities of 1) a space-time instantiation, 2) a dynamic-open system. 3) an indexical representational system. and 4) control aspects in space-time. These four qualities, if accepted in part or collectively, offer a perspective on why numeric identity retains a strong intuition in debates on persistence of identity, as well as indicate how a notion of qualitative identity may rather account for these intuitions. It is acknowledged that these qualities are not exhaustive and require further development, but nevertheless they are mentioned here to illustrate that certain qualities are consistent with certain intuitions about persons.

3.6 The person as a type with physical tokens

It is noted at the outset of this section that the type–token distinction, although retaining the general concept of the distinction, relates somewhat differently to the problem of persistence of personal identity than when as it related to the mind–body problem. Whereas in the mind–body problem (see section 2.5.2.1.) the tokens are different physical instantiations of the same mental type often occurring within the same person over different times (e.g., pain having different neural correlates at different times), within personal identity contexts, the person-type is an overarching unified concept related to the whole particular token body. Or, in other words, in terms of the mind–body focus of the previous chapter, the focus is on types with tokens that are part of the substrate (e.g., a particular neural pattern), whereas in terms of personal identity, the focus is on the substrate as a whole (the body) as it relates to persons. For those who assert that mind uploading is possible (e.g., Walker, 2014; Wiley, 2014), a person-type may be instantiated across multiple tokens (e.g., the original substrate and the upload substrate), whereas those who would deny that the upload is the same person (e.g., Pigliucci, 2014; Cappuccio, 2017) would assert that the person is inextricably connected to, or synonymous with, the original substrate (the original token).

A person-type may be assumed as referring to a particular person. I, the author of this thesis, may be Paul Type_1 and another person named Paul may be Paul Type_2, and so on. If each current person is a type, then it can be said that there is only one token person at present and the question, in relation to the MUP, is whether there is a possible scenario where two tokens may exist simultaneously and if so under what conditions?

Previously (section 1.5.5.), the current thesis presented a type sortal spectrum, where on one end there is a type of all physical phenomena and on the other an absolute unique type. This correlates the type–token distinction with uniqueness and multiple realizability, within a physicalist frame. What type of phenomenon is the particular person-type (e.g., Paul Type_1)? If the particular person-type is absolutely unique, then this view must face the task of explaining how it is that a physical phenomenon (assuming a physicalist stance) can be so unique (the uniqueness problem). Does each person have unique physical properties, or is the person a unique property? If so, it is a leap of faith (and against the empirical trajectory) that has no evidence in its favour at present. If the assertion is that the processes are unique, this raises further questions, such as how physical properties can lead to type of process that can only and in all possible worlds occur in one substrate. It may be considered that these processes are immensely complex and beyond the pragmatic reach of engineering, however, this would then be an engineering problem and not a metaphysical one.

Alternatively, it could be argued (e.g., Perry 1976) that the psychological continuity somehow lead to a 'further fact' (what Parfit terms the simple view of persons), where the substrate and the substrate processes collectively instantiate a unique person (PI = Substrate + Processes). The uniqueness problem repeats itself in questioning how a unique 'further fact' person can result from multiple realizable properties and multiple realizable processes. It appears that only the continuity of substrate (the biological solution) has a response to the uniqueness problem by focusing on the one property that is resistant to replication/multiple realizability, namely the property of occupying continuous space and time (numeric substrate identity). The current thesis has questioned why this one property should matter and has aligned with the view that it is what the substrate does that matters. If what the substrate does is what matters, and the properties and processes are replicable (multiple realizable), then it is unlikely that the person would be a 'further fact' but rather a replicable phenomenon. In relation to types, if all other types are multiple realizable, then it is likely that Paul Type $_1$ will also be multiple realizable (i.e., is unlikely to be a unique type).

Within the type sortal spectrum, the general-person type is one level of abstraction above the particular person-type. The general person-type includes all persons, including ourselves and everyone we know (here solipsism is assumed to be false). The general person-type has multiple tokens (e.g., the author and the reader) that share various qualities (e.g., cognitions, personality) that distinguish this type from other types (e.g., the type of metals does not share these qualities). Atoms, molecules, general-persons (the general person sortal), black-holes, and so on, are all types that have multiple tokens. Yet when it comes to the particular person type (e.g., Paul Type $_1$), some continue to assert that it is metaphysically unique (this claim is based on the notion that the person type is a numeric identity and that the primary condition is the space-time continuity of the substrate whole). Therefore, although possible, it appears that the insistence on the particular person type as having only one token is, at the very least, a divergence from other type levels.

The idea of uniqueness and whether it relies on metaphysical uniqueness or engineering uniqueness may be explored through a thought experiment. Imagine a genius, Dr Hatefrills, of unprecedented height, who is able to construct a machine so complex that for centuries no group of engineers is able to replicate it. In fact, there is even debate over what it is actually doing at times. Although there are some generic abstract descriptions, there is no clear understanding of how, or even what, it is doing. In this scenario, there is only one type and token correlation (type = token). Dr Hatefrills' machine is absolutely unique, one of a kind. Suppose that, sometime in the future, a group of engineers is able to fathom the machine so that functions are indistinguishable and they replicate the machine precisely. Now there are two tokens of the one type (type = token$_1$ and token$_2$). The MUP posits a similar situation, where evolution has designed a very complex

biological machine and one that we are currently unable to fully fathom. The appeal to historical correlation between one token and one type (the mind of particular person), therefore, may be an appeal to what has historically taken place but may not consider what could take place (metaphysics of possible worlds). The current uniqueness and inability to replicate may refer to difficulty of the engineering task for the MUP, but this appeal should not be used as a denial of the metaphysical possibility. As in the above example of Dr Hatefrills machine, uniqueness, on its own, may be a matter of complexity rather than a metaphysical assertion.

This section has not queried whether persons are types (as according to the type sortal spectrum an absolutely unique entity would be a type, albeit a type with only one token)[43] but whether there is the possibility of multiple tokens of the same type. The current section associates multiple realizability (the emphasis of the previous chapter) with the type–token distinction. That which is multiple realizable can be understood to be a type with multiple tokens, in that whatever is being discussed has multiple (more than one) instantiation (here being a physical token). Although types by convention (see section 1.5.5) allow for multiple tokens (e.g., words, letters, and so on), it would be philosophical sleight of hand to state that a) persons are types, b) types by definition allow for multiple instantiations (tokens), and c) therefore, that persons may have multiple tokens. Rather the current thesis allows for the possibility that types may be singularly instantiated (type = token) or multiply instantiated (type ≠ tokens).

Because the biological solution emphasises that the condition for persons are the space-time continuity of the substrate (a unique condition), the person (if the term type is applied to it) is unique and synonymous to the token. For each substrate (here described as a token), there can be only one person-type. Therefore, the biological solution may be formulated as the type = token (i.e., the substrate token is identified with the person type). That tokens have numeric unique conditions (substrate space-time continuity) is not queried by the current thesis and it accepts that a necessary (although not sufficient) condition of a token is that it occupies a particular unique space-time continuity. However, types do not have need of this condition (types may have multiple tokens that occupy multiple alternative space-time continuities). The matter, therefore, settles on whether persons are types that do not equal their token or whether they are tokens that are the unique types.

---

43 Furthermore, the issue of considering types as universals (e.g., Parfit, 1984, Olson, 2007) has not been addressed. This is because types as universals is only one avenue to consider types (Wetzel, 2018), and the type sortal spectrum may accept a view of types as universals or as abstract descriptions of particulars (i.e., the type cat does not necessitate a platonic, idealised cat and may simply include a descriptive category of all particular token cats that share cat qualities).

One way to explore the type to multiple token relation, would be to explore this from the physicalist stance, and this may proceed as follows:

(1) Persons are physical phenomena (the physicalist assertion).
(2) All known physical phenomena at the nomological level of our daily lives are multiple realizable (the assertion of the previous chapter).
(3) Therefore, persons are multiple realizable.
(4) What is multiple realizable is in principle replicable.
(5) Therefore, persons are replicable and may have multiple tokens.

As the physicalist stance is that all phenomena are in some sense physical, persons, as one of those phenomena, are deemed physical. The current thesis has emphasised that physical phenomena, in relation to the MUP, may either focus on the substrate or on what the substrate does. The previous chapter noted that what the substrate does (external and internal processes) has been identified by various physicalist stances with the mind (i.e., mind = substrate processes) and this has been the starting point of the current chapter. Furthermore, the previous chapter demonstrated the empirical trajectory is in favour of minds likely being multiple realizable/technologically replicable. If the empirical trajectory continues (i.e., we bet on the winning horse of science and technology), there will come a time when persons will be replicable.

What is multiple realizable is, in principle, from a metaphysical perspective, replicable. Put another way, if a phenomenon may occur at two space-time locations in two substrates (e.g. properties such as mass, or processes such as speed), the possibility that it may occur in some possible world through some form of technological artifact replication process is likely. Therefore, persons are likely to be technologically replicable. The biological solution discussed in the current thesis has largely affirmed the first premise (that persons are physical) but denied the second (that persons are multiple realizable). The primary method, for the biological solution presented here, has been to make the claim that the person *is* the substrate (in some way). To do this, proponents of the biological solution have used thought experiments aimed to illicit the intuition that the process of identification of persons is to 'follow the body' and that there can only be one of the same person (numeric identity). This leads to what the current thesis terms the 'token only' assertion (type = token). If persons are to hold numeric identity and persons are physically instantiated, then it follows that the physical token is the object of identification.

Olson (2007) discusses person-types by aligning the type–token distinction of stories (Olson using that of *Moby Dick*) and that of computer programs. Olson acknowledges that if a person

were to be a type similar to novels or programs, that multiple tokens would be a possible outcome. However, Olson is not convinced and points to "grave problems" (ibid., p. 148) that may be simplified as:

(1) Persons are numeric entities that "do things".
(2) Types/universal are not numeric entities that "do things".
(3) Therefore, persons are not types and, therefore, may not have multiple tokens.

As has been discussed previously 1) can be denied, as persons may be qualitative processes and not numeric entities. Premise 2) may be contested either on the grounds that types do "do things" (i.e., those who hold that universal types are causal) or on denying that types need to be universals as is the stance of the current thesis. This is a particular problem for Olson (ibid.), who is here attempting to critique Dennett (1991), however, for Dennett, the 'heavy lifting' (i.e., the causal work) is done by the token hardware (see section 2.5.2.4. on Dennett's intentional stance). That the software program is not 'doing things' but the hardware token is, does not negate the software being a type with its own nomological boundary (although it is causally reductive). Because 1) and 2) may be contested, so may the conclusion. A similar line of reasoning is presented from critics of the MUP (e.g., Hauskeller, 2012; Cappuccio, 2017), where persons are attributed to types that are numeric things (the category mistake of conflating the substrate with what the substrate does) and, from there, that persons are numeric tokens (type = token). From this perspective, if the person is the token (numeric), the MUP is not feasible as any replication would lead to a new person (a new token). However, if an alternative premise is made (i.e., that persons are not numeric things but processes that may have multiple instantiations), the arguments fall short.

Williams (1973), a proponent of the biological solution, discusses person-types in relation to the reduplication problem (what the current thesis calls replication and would not describe as a problem). Williams puts forward a scenario where information contained in the brain could be placed in alternative substrates, which could then "print off more than one person in accordance with these conditions" (ibid., p. 80). The original is viewed as a prototype person with various versions (tokens) being created. For Williams (ibid.), space-time instantiations become important, in that each token-person would diverge once interacting in a new space and time (see section 3.7).

Williams considers a particular person named Mary Smith and asks whether one loves the token-person or the type-person. If there were multiple tokens of Mary Smith, would someone who loves her (the person-type) cease to love her if the Mary Smith token (the substrate) varies? This love is likened by Williams to loving a work of art (his example of *Figaro*), and that it may be

acknowledged that though there are variations in performance (tokens) that one may love the type and accept a poor performance (token) over no performance at all. This leads Williams, in his chapter titled *Are persons bodies?,* to some cautious conclusions. Loving a person-type may vary on the token (such as one preferring one performance of *Figaro* over another) and he claims that we are not to "undervalue the deeply body based-situation" (ibid.). However, to love a person is not to love a body and to say so "is more grotesquely misleading than it is a deep metaphysical error" (ibid.). So, for Williams, the person-type is something more than the person-token yet we are not to 'undervalue' the token. Williams acknowledges that this notion of person-type requires more work, but still leans toward the notion that the person-type is the same as the person token (types = tokens).

Parfit (1984) comments on Williams's Mary Smith and the type–token distinction. Parfit considers a person loving Mary Smith and that the original is destroyed and a replica made. Will the replica be loved? If so, then what is loved is a type and not the token. Parfit now imagines two scenarios, one of branching and one of non-branching. The non-branching scenario is discussed here with the branching scenario being discussed in the next section 3.7.

In the non-branching scenario, the replica is seen as a youth-preserver, where Mary, in an attempt to stay young, enters the replicator at age 30 once a year and a replica (biologically at the beginning of age 30 phase) exits. Parfit imagines a person falling in love with Mary and asks what the response should be after she has entered the replicator and the younger replica emerges. One can either respond with grief (if one believes Mary has in fact died) or love (if one believes Mary continues as the new substrate). For Parfit (ibid.), if the psychology continues, the person continues and, therefore, the replica is most likely to be loved. Loving a person, is viewed by Parfit as a process that involves shared memories. As the replica will have these memories (albeit a Q-memory), so the shared history of loving may continue. But, what of loving the particular token (the substrate)? Parfit accepts that this is one type of love (or possibly lust) but that this is a shallow form of love and not the form of love that he takes to be important to the discussion. For Parfit then, it is the R-relations[44] that matter, and what is loved, irrespective of the substrate, is that within which the R-relation is instantiated and it is this scenario that is, then, defined as the type≠token.

44 Parfit (1984) associates types with universals and then proceeds to deny that persons are universals but rather R-relations that may be instantiated in multiple substrates. Because the current thesis uses the type sortal spectrum (which does not necessitate that types are universals), Parfit's view is akin to types ≠ tokens, where the type is the R-relation and the token is the alternative substrate the R-relation is instantiated in.

The current thesis uses the terms type and token similar to Walker (2014), who uses the example of the play *Hamlet* as a type (the *Hamlet* type) that can be instantiated in multiple tokens. The thrust of his argument is to show how, by applying the type–token distinction, it allows for branching (to be discussed below in section 3.7). To summarise Walker's (ibid.) general argument, his claim is that identity means the preservation of a phenomenon. Furthermore, types are about qualitative identity and tokens are about numeric identity. The *Hamlet* play identity is a type and the type may be preserved despite different tokens being destroyed (so long as the play type of *Hamlet* continues in some token). For Walker (ibid.), personal identity is a type, like a play identity is a type, and although he acknowledges the contested ontological status of abstract descriptions such as types, he nevertheless affirms that what holds true for the play type should hold true for the person type. Therefore, personal identity may be preserved across tokens. One possible counter argument, from the biological solution, that Walker does consider is presented as follows:

> "P1 [Premise 1]: Multiple replicas X, Y, Z ... of an individual O (the original) are numerically [token] non-identical with each other.
> P2 [Premise 2]: Preservation of personal identity requires preservation of numerical [token] identity.
> C [Conclusion]: Therefore, not all replicas X, Y, Z ... preserve personal identity of O" (Walker 2014, p. 173).

Walker's (ibid.) response to this argument is that it is "question begging" (p. 174) and that it "prejudges the issue" (ibid, p. 174), as well as to query the need for numeric identity (P2) as arbitrary. Walker's criticism, therefore, aligns with the current thesis's criticism of the numeric identity, which may still be a coherent assertion (there is nothing incoherent about asserting that persons are their numeric token substrates), but that the assertion itself requires further validation.

Walker (2014) allows for persons to be sentimental about tokens and that this may explain some of the desire to retain the current token (the biological substrate). For example, the first token book purchased of a particular book type (e.g., *Moby Dick*) may have sentimental value and we can imagine being upset if that particular token book were to be lost. However, given the choice of not having that particular token and not having any token of the same book (the type), persons may likely opt to at least have the second token (and retain the type *Moby Dick*). Sentimental value is, however, arbitrary and may vary according to person and community. For example, a trans-humanist community may value the upload substrate *more* than the original biological substrate (e.g., a

synthetic substrate may be less susceptible to illness). If the value of a token is based on the evaluation of the context, it is not an intrinsic quality that matters and, therefore, is not a sufficient and necessary metaphysical condition for personal continuity (although it may play a role in what is decided by particular persons).

If one assumes the psychological solution, the type–token distinction may be explored through various means depending on preferred stances as to how the mind is understood. The current thesis now presents one such argument from the premise of minds as representational systems (see section 2.6.2) to persons being type to multiple token applicable. The argument may be formulated as follows:

(1) All representations are one-to-many type–token distinct.
(2) The mind is a representation system.
(3) Persons are either a) the mind, b) partly the mind, or c) not the mind at all.
(4) Therefore, persons, if a) is accepted, are one-to-many type–tokens; if b) is accepted, are partly one-to-many type–tokens; or c), are not one-to-many type–token distinct.

The first premise (1) may be based in the notion of type–tokens and its use as such in the original work of Peirce (1906) that reflects the type sortal spectrum. The idea can be expanded to include all representational systems currently known such as modern computers, numbers, and so on. In each example of types and tokens, where representations are used, there is a one-to-many relation between the type and the tokens. The second premise (2) is the assertion of the representational view of the mind, which within this particular argument is assumed to be true. These two premises then lead to the three options as presented throughout the current chapter of a) PI = Processes where the mind is representational processes, b) PI = Substrate + Process, where the mind is partially a representational process, and c) PI = Substrate, where the mind is not the representational processes but the substrate that performs these processes. If persons are minds, and minds are representational systems, then persons are type to multiple token appropriate. If persons are, in part, minds (option b), then this part is type to multiple token appropriate. It then needs to be established (outside of this representational argument) whether the other part is also type to many token appropriate. If persons are not minds, then the argument is irrelevant.

This section has shown that within the type–token distinction, persons (if viewed as types) may, in principle, have multiple tokens. This is presented as the relation type≠token, in that the type may be instantiated across multiple tokens. This type≠token relation, if appropriate for persons, would allow for replication of persons across substrates. Alternatively, if the person as a type to

many tokens is denied (here determined as the token only view or the type = token view), then there can only ever be one metaphysical numeric person bound to a particular substrate and the MUP is not feasible (any replication whether in part or whole would be a replicant that does not retain the identity). Furthermore, what this section has shown, is that when discussing types and tokens, if one removes the numeric identity and persons as entities assumptions, the rest of the arguments have little traction. Perhaps the same can be said of arguments in favour of the psychological solution, but this would, at the least, lead to an intuition stalemate in relation to the type–token distinction.

3.7 Branching identity

The idea of branching is not a unique idea to personal identity and occurs across multiple domains and, in general, can be defined as a divergence from a common ancestor to more than one offshoot. Consider a tree a trunk with branches that emerge from this trunk, which, in turn, have further branches emerging from these branches, and so on. The idea of branching can be seen in theories of evolution (e.g., Dawkins, 2011, Dennett, 2013), according to which there are various branches of species (sometimes called the tree of life). Each species is said to branch off from a common ancestor and is said to be a different species when reproduction can no longer occur across these groups of animals. Mitosis is another biological example of branching in biology. In mitosis, the cell divides and produces two numerically separate cells with the same singular cell ancestry. Consider monozygotic twins who, at some stage in their personal ancestry, share the same space and time substrate *in utero*. As the foetus's develop, they branch and two separate substrates emerge, each with their own biology (epigenetic activations and so on) and with different interactions with their environment (they occupy different space and time and, therefore, have different experiences). Now, if branching can occur within other domains, it is at the least possible that it could be the case for persons.

The notion of branching takes seriously the notion that persons are processes that have space-time instantiations as well as that there is the possibility of one-to-many substrate instantiations of these processes. The general idea is that a singular person may branch into two (each having equal claim to the ancestry of the singular person), but at the time of branching (e.g., fission), each process diverges into new persons. Consider a person who is instantiated with a particular original substrate, such as is currently the case with all humanity (Person O). Currently there is no branching as each person process continues to be only instantiated in one numeric substrate. In the branching scenario, at some point, this person process branches into instantiations in two new substrates and becomes two separate persons (Person A and Person B). Assume that these two new substrates are

not the original. Person A is a different person to Person B as they each have divergent experiences, each interacting with the environment in space-time (the nature of space-time instantiations of dynamic-open systems). For example, let it be assumed their employer can only pay one salary and that one salary is insufficient to support both substrates' needs. Person A keeps the job and Person B gets another one. Now, in two separate work places, each person makes new friends and has new experiences, which in turn, impacts on the personal development of each and, therefore, each of Person A and B becomes a different person (this type of the thinking is common in the counterfactual thinking we do in our own lives and statements such as 'I would be a different person if I had taken that job/married that person/moved overseas', and so on). Both persons have claim to Person O as the common ancestor person yet, over time, Person A and Person B diverge (branch) into different persons.

Wiley (2014) puts forward an interesting thought experiment, where the two persons, after branching, are placed in two rooms that are qualitatively the same (the two rooms occupy alternative space-time locations but, aside from this, are precisely the same). These two persons are then given the same communication from the outside world (e.g., recording of loved ones and so on) and live their lives in precisely the same way. In this scenario, we have two persons who are having the same qualitative interactions with their environment. It, therefore, raises interesting questions as to whether they have actually branched as persons, as 'persons' are defined here as the same qualitative process (while acknowledging that they have branched numerically). Within the current section, the thesis focuses on the concept of two separate substrates with a continued process that occur in two separate environments.

Branching is developed within Parfit's (1984) evaluation of the Mary Smith replica scenario (originally in Williams, 1973 and mentioned in the previous section 3.6 on type–tokens), which he expands from a non-branching scenario discussed above to include a branching scenario. Parfit allows for multiple tokens to exist simultaneously (previously in the non-branching scenario; as each new substrate is made, the previous substrate was destroyed). In Parfit's branching view, there are multiple tokens of the same person-type such as multiple Mary Smiths (later Parfit actually switches his example to the actress Greta Garbo, but the idea of multiple tokens that branch off from an original token is the same). Would these be the same persons (person-type) or would the development of each token in a new environment with different interactions lead to different person-type Mary Smiths? Parfit (1984, p. 302) talks of 'successive selves' and an 'ancestral self', where each successive self may be distinct over time but have the same ancestral self. In branching, because each successive self has the qualities of control, space-time instantiations, and is indexical

(i.e. develops divergent R-relations), the argument is that each new substrate will lead to the emergence of a new person, yet each will have claim to the same original R-relation that was previously only continued in one token substrate. For Parfit, what matters is that there is a continuity of psychology (the R-relation), and if a person may trace back this process (irrespective of the substrate that currently instantiates that process), then in this sense it is the same person.

Those (e.g., Wiggins, 1967; Williams, 1973) who hold to the biological solution are faced with difficulties when thinking of fission cases. Williams (1973) states that claiming two persons being the same person is "absurd" (p. 19) and that it breaches traditional space-time continuity of persons as he "cannot see how it can satisfy the logical requirements of being a criterion of identity" (ibid., p. 19). Furthermore, he presents what he calls the reduplication problem (where the idea of a person being replicated is viewed as a problem). What underlies these 'problems' is the assertion that there cannot be two identical persons occupying different space-time (i.e., personal identity is numeric identity at the nomological boundary of bodies). Williams (ibid.) acknowledges that there may be problems for his own views in such cases as fission (here Williams imagines an amoeba-man who splits in two), but that this is not the case for persons at present.

That persons currently coincide with a particular substrate is not contested and it should be noted that, at present, there is no problem of personal identity in the sense of everyday attribution (folk PI works in the current state of affairs). The problem comes into being only in so-called problem cases, of which the MUP is one such case (also see section 3.4). From the perspective of the current thesis, it appears that Williams has conflated space-time instantiations with space-time continuity of a particular substrate, as well as conflating processes with entities (entities have numeric qualities whereas processes need not). The same problems are mentioned within Wiggins's (1967) case, where fission is likely to allow for branching possibilities. The problem, for Wiggins, is that there can only be one of 'you' and Wiggins minimises the problem of branching (fission) by reaffirming that persons are substances (his substance sortal emphasis) and, therefore, that there can only be one person (a non-branching assumption based on numeric identity). The so-called problem of branching is, therefore, only a problem to those who hold to the dual assumptions of the biological solution that a) there needs to be numeric identity to personal continuity, and b), that persons need to be things (substrates/substances). If one holds to the views of the current thesis (that persons are continuations of qualitative processes), then the branching of persons is no more absurd than a domino cascade, software programs, and a book having more than one token.

For those who hold to the closest continuer theory (Nozick, 1981), which includes the numeric identity assumption, further difficulties emerge. Imagine two persons who have brain-bisected fissioned from one original substrate instantiation (person O), with one fissioned person

(person A) retaining one more neuron than the other (person B). This single neuron becomes the difference between continued personal identity (person A is the same as person O) and the replica as a new person (person B is a new person not A or O). This single neuron (which may die off at any time or may perform no function) is the difference between one fission retaining the original identity and being a different identity. Now imagine that neither person is distinguishably different (such as in the teleporter), in this instance neither person O, A, or B retain the personal identity (O has ceased to be and A and B are new persons and neither is closer). The insistence on non-branching, therefore, creates difficulties that lead to personal identity being arbitrarily assigned. If branching is allowed, then both persons have equal claim to the ancestral person O, as well as becoming two different persons as they diverge from each other (the result of qualitative conditions of persons being space-time instantiations, dynamic-open, indexical, control systems).

If persons are the processes and not the substrate that instantiations these processes (qualitative identity, type ≠ token), then these processes can, in principle, continue across substrates (the necessary and sufficient condition for the MUP). What can continue across one substrate could, in principle, continue across multiple substrates. From this it can be inferred that the person may branch. Certain practical limitations (conditions) of branching are discussed here. For example, if persons are deemed to be dynamic (change over time), what would be a sufficient time to pass for there to be change? Imagine a machine-learning program that learns through interacting with the environment. The program is stopped at some point and, at a later time, is restarted. In the time that lapses between the program being stopped and restarted, there is no process instantiation within the machine (the program is in stasis), and so this time frame is incidental to the learning process (the machine could be stopped for a day, a month, or years, and it makes no difference to the learning process). However, once the machine is restarted, the process will take some time before adapting to the new environment (the process will continue and the program will be different from its past instantiation).

For the person as process there may be various time frames presented as to what may be relevant to changes in the cognitive processes (here assuming the mind as the necessary and sufficient condition). Cerullo (2015) suggests 30 milliseconds (based in part on the work of Efron, 1970), whereas others (e.g., Matthews, Stewart, and Wearden, 2011) may prefer 50 ms and add further dynamics such as intensity of stimuli. For the sake of argument, the current thesis has taken 30 ms (the lower time frame) to be used within this section. The precise conditions may vary and will rely on further experimental studies, but what is being emphasised here is that any dynamic process that is instantiated in space-time may have a threshold of sorts that determines when the process may be said to be the same and when it may be said to have changed from a previous

instantiation of the process. Returning to the example of the machine learning program, it is further noted that the program could be instantiated in an alternative substrate (e.g., another hardware substrate) and that there could be multiple instantiations. Within the multiple instantiation scenario, let it be assumed that the program was copied while in the stasis phase. At this point there are multiple qualitatively identical programs and so they are the same program. It is only once each of these is restarted in a different environment that each instantiation becomes different (branches) from their ancestral program.

Assume that persons are processes (what the substrate does). On this stance, processes are physical processes (however defined) and require a substrate and interactions at the appropriate nomological level. Because persons are deemed to be processes, changes in substrate (e.g., the replacement of molecules or the replacement of the entire substrate) are deemed non-essential to the continuity of persons. The claim that the changes in the substrate are non-essential, does not negate supervenience or any other physicalist notion that captures that the processes are in some way dependent on the substrates. Rather the emphasis on process helps to establish which aspects of the substrate properties need to be retained. For example, consider the two slinky substrates (one metal and one plastic) joined such that the wave function may continue across these substrates. It does not matter that one slinky substrate is plastic and the other is metal, nor that these substrates may occupy distinct space-time continuity, but should the substrate change drastically (e.g., a slinky were connected to a brick), the wave function would cease to continue (e.g., in the brick substrate the wave function would terminate). The properties of the substrates, therefore, matter. Furthermore, consider the example of supervenience, where an image supervenes on the micro-properties of colour and their arrangement. These micro-properties may not be changed if the image is to supervene (i.e., the properties of the colours and how they are arranged needs to occur in the substrate). However, whether these micro-properties are constituted from pixels or paints or whether the image supervenes in one location or another is not essential, as the image may supervene irrespective of these substrate properties. Therefore, it is acknowledged that change in the substrate may impact on the process (e.g., wave function or image) but so long as the process is maintained (irrespective of substrate) that this is what matters. The emphasis here is, therefore, on the continuity of processes (which do not require the same substrate and allows for discontinuity of space-time) as the necessary and sufficient condition for the continuity of persons.

But, although accepting branching may resolve many of the problems of personal identity offering a coherent perspective on the intrinsic qualities that explain the metaphysics of persons, this is not to say that it is without problems. Some of the implications of branching may be presented with the following options:

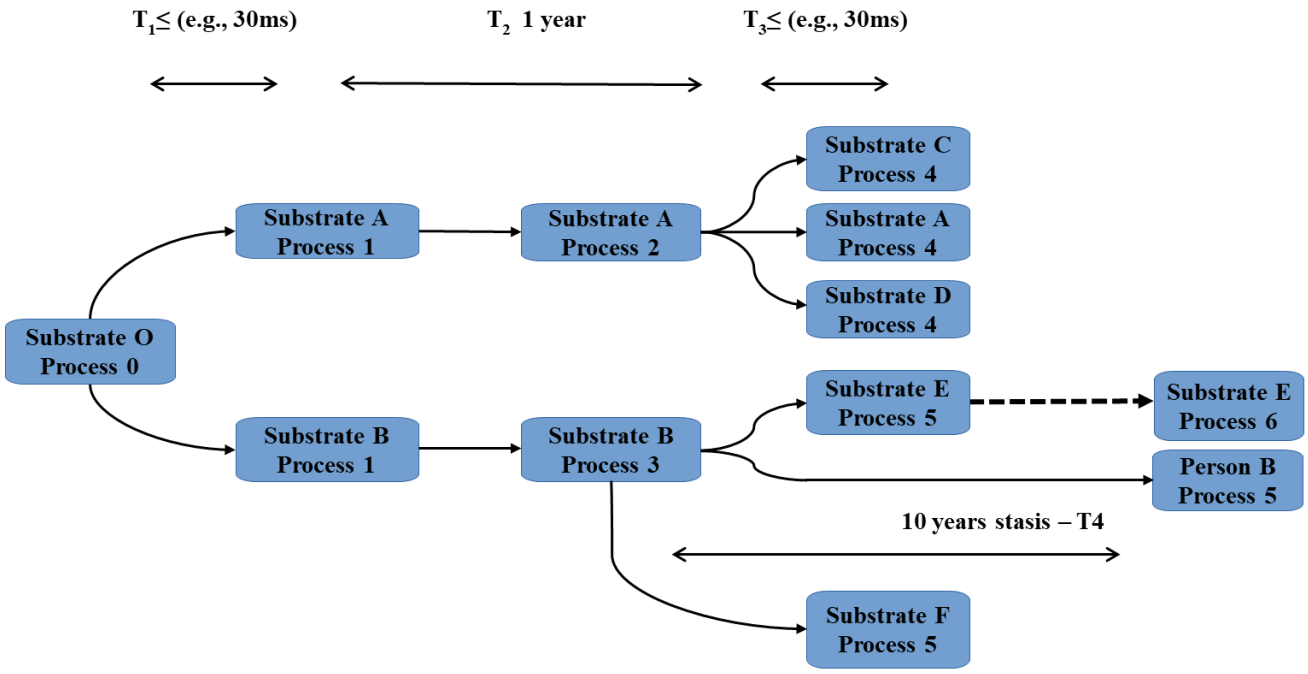

Figure 3-2: Branching implications

On the left, there is a particular substrate (Substrate O) and a process (Processes 0). As it is acknowledged that persons are dynamic processes, this process is a continuity of previous processes from the processes of the child, to the teenager, to the younger adult, to the person at a particular time. Each event in time may be seen as an adaptive change of the process and, therefore, the process has had many iterations throughout the life of the person thus far. Process 0 is simply the designator within the diagram of the process at the outset of this thought experiment. Prior to process 0, it is acknowledged that there would have been multiple processes leading up to this time. At T_1, the process is continued in two alternative substrates (Substrate A and Substrate B). Assume that Substrate O has ceased to be for some reason (e.g., natural death of the substrate and the resulting cessation of the processes). At a particular time (T_1), the substrates A and substrate B have not had sufficient time (here assuming a time less than 30 ms) to diverge in any possible way and, therefore, at this time, the process has all the dispositions, properties, and whatever else is deemed necessary for the process to continue, which is the result of Substrate O's history of interactions with the environment and development. At T_1, Substrate A and Substrate B instantiate the same process (Process 1). Therefore, there is a type-person with two token-persons. As there is no qualitative difference between Process 1 of Substrate A and Process 1 of Substrate B, and the process is deemed the necessary and sufficient condition for persons, Substrate A and Substrate B are said to instantiate the same person (Process 1) at T_1. At T_2, a year has passed and substrate A (having

different interactions both internally and externally) has developed into Process 2, while Substrate B's Process 1 has developed into Process 3. This is the inherent nature of persons and the basis of the persistence problem, in that persons change over time based on interactions. This is the same dynamic process of change that you or I, with our singular biological substrate (our bodies), has experienced in life, which is the continuity of the process as adaptation and emergence of new states of self (process iterations).

Now, at $T_{3,}$ there is a further branching. Substrate A's process is replicated in two further substrates (Substrate C and D). At this time, Substrate A continues to exist and now there are three substrates A, C, and D with the same process (Process 4). $T_{3,}$ like $T_{1,}$ is also sufficiently limited (30 ms) such that at T_3 there is no qualitative difference between processes and, therefore, it can be said that the same person (Process 4) occurs simultaneously in three substrates. As time continues, Process 4 initially instantiated in different substrates in space and time, will develop into three further processes (persons). Therefore, within the time frames of our everyday life there are not two persons occupying the same space-time, however within certain limited time frames two persons (processes) may be determined as occupying different space-time locations and space-time substrates at the same time (e.g., at T_1 and T_3). This would mean that numeric identity is non-necessary for the preservation of persons, although certain numeric qualities (e.g., <30ms, space-time instantiations with a physical substrate, control, and so on) are.

Now consider a further possibility in relation to substrate B, Process 3, where a similar branching occurs but the one substrate's processes are not activated. Substrate E and F are created and instantiate Process 5 and at the limited time (T_3) they are the same person (Process 5). However, as substrate B is not active (the process is in stasis), it cannot be said to be Process 5 at T_3. At T_4 however, Substrate B is activated and for a limited time instantiates Process 5. Therefore, we have Process 5 instantiated not only in two separate substrates (Substrate E and Substrate F) but at two different times (T_3 and T_4). It is further noted that at T_4 Substrate E and Substrate B are different processes as substrate E has had the time to further interact and develop into Process 6, whereas Substrate B is only instantiating Process 5 (which is in Substrate E's past at T_3).

In the branching perspective, where persons are processes, there can be the identity of an ancestry of person processes. For example, all of the processes share a common ancestor of Process 0, however, Process 4 does not share any ancestry with Process 3 (similarly, Processes 5 and 6 share no ancestry with Process 2). The identity claim (that a phenomenon is the same person) is, therefore, a sortal/definition/description, where a current process may trace back to an earlier process. This process can be stopped and restarted, move across substrates, be placed in stasis, and branch into multiple divergent processes. Currently, in human history, there is only one continuity

of the process. However, given that physical processes (and properties) are multiple realizable, it can be inferred that persons (as a physical process) could, in fact, have multiple avenues for continuation (branching).

It is not difficult to see why many persons may want to reject the branching stance. For example, persons may see it as too complex and prefer the simplicity of the numeric identity and the biological solution. Furthermore, the idea that there could, potentially (even if only for a limited time), be two 'you's' in two different locations seems counter-intuitive to common experience. Not only is branching complex, but it would also have moral and pragmatic implications, such as who gets the car? The current thesis accepts that, as with any complex issue, there are multiple difficulties and here makes no effort to resolve all of these. Rather, the thesis has presented that persons are likely processes, that the processes that matter to the continuity of persons are likely psychological processes, and that, if this is the case, there is the possibility of branching. If the uploader (due to whatever reason) wishes to only have one process active at a time, then they are free to do so, however, it is acknowledged that the choice to do something does not negate the metaphysical possibility of an alternative choice (i.e., choosing to have multiple branched persons).

3.8 Chapter conclusion

The current chapter has given a very brief overview of the complex field of personal identity in relation to the persistence problem. Relevant themes have been emphasised and discussed in the context of the MUP. In the previous chapter, the thesis evaluated the mind–body problem and determined one of the primary obstacles to solving the mind–body problem lies in the category mistake of conflating entities (things) with processes (how they hang together), and this is largely based on assumptions of the manifest image as opposed to the scientific image (which is deemed by the current thesis to be the preferred epistemology).

In relation to the subject of persistence of persons, the current chapter has echoed this approach and taken the stance that persons are minds (the psychological stance) and that minds are processes. Processes relate to qualitative conditions, in that they have a one-to-many relation (R-Relation or non-numeric 1:1 relation) with each other. The current thesis has emphasised some of these qualitative processes that are the necessary and sufficient conditions for persons that may account for folk intuitions in relation to persons relating to numeric (non-branching) identity claims. These qualitative conditions include (although are not limited to) space-time instantiations, control, representations, indexical orientation, and dynamic-open conditions.

In essence, the conclusion of this chapter is that, should the persistence of persons require a particular substrate, then the MUP is not feasible, however, should the persistence of persons

require processes (at whatever nomological level) that could, in principle, be instantiated in alternative substrates, then the MUP is feasible.

4 Integration for the MUP

Having explored the metaphysics of minds and persons in relation to the MUP, the current thesis now turns to integrate salient themes that have emerged. To do this, the chapter presents three sections.

First, section 4.1 overviews the terrain already covered, aiming to establish a conceptual map of what the necessary and sufficient conditions for the MUP could be. Two tables are produced for each of chapters 2 and 3 with attention to how each philosophical distinction (the mind–body problem and the problem of persistence of identity) impacts on the metaphysics in relation to the MUP, as well as constraints that each view places on potential upload avenues (the MUP matrix). This is the core of the thesis as it reflects the primary problem of what the necessary and sufficient metaphysical conditions for the MUP would be. The section reaffirms that if multiple realizable physicalism in relation to the mind–body problem and the psychological solution to the persistence problem are retained, then the MUP is metaphysically feasible.

Second, section 4.2 presents the golden thread that can combine the suggested solutions to the mind–body problem and the persistence problem (as summarised in the previous section), namely the current thesis' emphasis on process. Although process as a means to understand the mind and persons is not a unique concept (see previous chapters), the application as a unifying concept within the MUP in relation to the mind–body problem and the persistence problem is. In this section, it is argued as to why processes should be the primary emphasis and how this impacts on the MUP and various themes that have been developed throughout the thesis. Processes of mind and persons (see previous chapters) are categorised under process-selves. Within this view of process-selves, the processes of minds and persons are deemed to be the same processes (identity claim) and, furthermore, that properties are distinguished from processes and it is only processes (not properties) that are selves. A flow diagram here integrates the main themes of the thesis. This section is a more refined view of what the metaphysical nature of selves (minds and persons) is and allows for the feasibility of the MUP.

Third, section 4.3 presents a rough sketch of the thesis' own preferences of selves (persons and mind) through a novel concept of the self/mind, the efferent-self. This builds from the previous sections and aims to further develop a more specific form that process-selves may take in identifying efferent processes (to be clarified in this section) as the intrinsic metaphysical processes that are selves.

4.1 Metaphysics – necessary and sufficient conditions and constraints

What has been demonstrated in the thesis, is that the two necessary and sufficient conditions for the MUP are that of multiple realizable physicalism and psychological continuity. Without these two premises, the feasibility of an artifact that can instantiate a person across substrates becomes less and less likely. In this section, each of these premises are explored in relation to literature discussed within the respective chapters.

The reader is reminded of Sellars' (1963) distinction between the manifest and scientific image and that the current thesis has preferred the scientific image as its epistemology. Furthermore, the thesis has adopted Sellars' (ibid.) philosophical framework for distinguishing between things (entities/substrates) and how they hang together (processes), as well as adapting his views to include the possibility that, at times, things hang together to form new things (weak or strong emergence). It is deemed a category mistake to confuse entities (things) for processes (how they hang to together).

A novel concept of nomological boundaries, as a way of determining what domains of scientific study are being considered in relation to each other for the establishment of necessary and sufficient conditions for metaphysical concepts, was also presented throughout the current thesis. In philosophy of mind (whether the mind–body problem or the persistence problem focused on in the current thesis), the primary nomological areas of interest are between the psychological/mind and that of the biological/bodies. Although the current thesis has expanded (e.g., included behaviours) and made further distinctions (e.g. higher level and lower level processes), the general distinction between a thing/entity (i.e., the substrate), and a process (i.e., what the substrate does) has been maintained across chapters. The substrate and the substrate processes concept will be further developed in sections 4.2 and 4.3, whereas the current section focuses on summarising the arguments for multiple realizable physicalism and psychological solutions to the problem of persistence of personal identity as the necessary and sufficient conditions for the MUP.

4.1.1 Multiple realizable physicalism, the mind-body problem, and the MUP

The mind–body problem and the metaphysical nature of the mind are represented in Table 4-1. Distinctions are presented in the table following the format of chapter 2 and the categories presented with each distinction (briefly discussed in turn) relates to the columns of feasibility in relation to the MUP and constraints. The constraints column is further divided into metaphysical and engineering constraints, each placed on a spectrum from less to more likely (the binary columns of less and more are illustrative to emphasis the two ends of the spectrum). Although the metaphysical

constraint is the primary focus of this section, the table includes a spectrum of engineering constraints as well. This is to account for some of the critiques within the literature (e.g., Pigliucci, 2014, Searle, 1980, 1990, 2004) that focus on constraints of current technologies. The aim here is to acknowledge that certain engineering constraints are present, while affirming the MUP as metaphysically feasible.

Within chapter 2, the kind of substrates at issue in the mind–body problem context has largely been discussed within the ABC options of (A) artificial, (B) biological, and (C) contraindicated substrates. The category of artificial substrate is less constrained in terms of the MUP project as there are already many complex artifacts of a similar nature (e.g., computers as representational systems), although it is acknowledged that, at present, these are insufficient for the MUP agenda. The biological category is more constrained in terms of the MUP project, in that there is currently less knowledge and artifacts that would be considered similar to processes of the mind (e.g., serotonin re-uptake and synaptic chemical reactions). Simply put, the more processes an artifact needs to perform and the greater the amount of properties, the greater the difficulty of (re-)creating that artifact. Furthermore, the thesis has introduced and developed the term 'nomological boundary' to orientate where each theory has placed its preferred level of scientific analysis necessary to determine whether an artifact can emulate/replicate the desired phenomenon. The more nomological boundaries needed to be integrated, the greater the engineering burden on the MUP project.

The metaphysical constraints also express a spectrum of less to more likely in relation to the MUP. Similar to the engineering constraints, a metaphysics of mind that has more properties, such as including both standard physical and further physical properties, increases the amount of work that needs to be included in the metaphysical theory. It should be noted that constraints in relation to the MUP makes no assertion as to the validity of the metaphysical claim. A less constrained metaphysics for the MUP, such as standard physical property stances (less constrained because there is no need to include further physical properties) may not be valid, yet it nevertheless offers fewer constraints to the MUP project.

What Table 4-1 illustrates, is that for the MUP to be metaphysically feasible, any number of stances in terms of the mind–body problem context may be retained, so long as the conditions of physicalism (the broad notion of physicalism in asserting that the substrate is physical) and multiple realizability are upheld. Hence a necessary and sufficient metaphysical condition for the MUP is that of multiple realizable physicalism. The rest of this section presents an explanation of each row of the table, having considered feasibility of the MUP in general.

The non-physical substrate row (within this thesis associated with substance dualism) is allocated to the non-feasibility of the MUP column. As such, no constraints are allocated, as the entire project of the MUP would not be feasible. In essence, the view here is, that should the mind be constituted by properties and processes that are beyond the purview of science (whether current or some future science), then the possibility of an engineered artifact that could produce the mind of a particular person is non-feasible. This non-feasibility of replication means that the properties and processes are non-multiple realizable and that which cannot be realized across a different substrate would nullify the MUP (see section 2.4).

In contrast, the physical substrate rows (further rows are presented as further distinctions within this overarching category) are feasible with less or more constraints. This assertion rests on the arguments of chapter 2, which demonstrated the probability of replication based on multiple realizable physicalism; that physical properties and processes, from different physicalist stances that include both standard physical or further physical substrates, are likely to be replicable based on the empirical trajectory of known physical properties and processes at the relevant nomological boundary. The thesis has followed the philosophical tradition of focusing on the primary nomological boundaries between bodies, minds, and the environment and these act as the further distinctions within the overarching physical substrate category (in the table these are classified as external and internal processes as discussed in chapter 2 and to be further discussed below). As all physical substrate rows are multiple realizable and allocated as such within Table 4-1, they reflect the metaphysical feasibility of the MUP.[45] In essence, if physical properties or processes can be replicated, they are multiple realizable and this replication would allow for the feasibility of the MUP as the replication could be said to be the same as demonstrated in chapter 2.

The next distinction within the category of physical substrates is that of standard physical processes, which are, in turn, further categorised under external and internal processes. External processes are further categorised as behaviourism and externalism/extended mind. In the category of behaviourism, it is claimed that it is necessary and sufficient for the behaviour of the substrate to be replicated for it to be deemed the same mind. This category is indicated here in the table as

45 It has been acknowledged, within chapter 2, that there are physicalist stances that would deny the multiple realizability necessary for the MUP. For example, the type identity theorist who has a flat realization of the mind as only being realized in a particular substrate. These have been discussed in chapter 2 (see section 2.5.2.1), deemed unlikely by the current thesis, and do not form part of the current chapters aim of integration. To clarify, not all identity theorists need assert that the mind is multiple realizable, but those who would (e.g., Smart, 2017; Akand, 2018) are the form of identity theory emphasized in multiple realizable physicalism as related to the MUP.

being a less metaphysically constrained stance as well as offering less engineering constraints. The nomological boundary of behaviour places the necessary and sufficient conditions in the context of the observable external processes of the substrate. Should an artifact behave in the same manner (produce the same behaviour), it is the same mind, and therefore, should this be achieved, the MUP would be successful. The engineering constraints are also fewer, as both biological and artificial substrates (A and B options) could in principle perform the same behaviours (see sections 2.5.1.1 to 2.5.1.4.).

Although the category of externalism/extended mind includes internal processes, this thesis has categorised these processes under external processes as this is the emphasis that is being explored. Furthermore, it is possible that an externalist/extended mind stance may be a property dualist (i.e., hold that external processes are necessary, but that these would need to interact with internal further physical mental properties to produce behaviour). The current thesis, while accepting this possibility, has chosen, for the sake of brevity, to allocate these stances as standard physical properties. Should this alternative option of an externalist/extended mind property dualist be taken, this alternative view could then be explored in conjunction with the further physical properties presented below. Here the standard physical properties are accepted with the emphasis on external processes and, therefore, what applies to the category of behaviourism, applies to the category of externalism/extended mind in so much as they perform the same external processes. However, the standard physical property externalism/extended mind category presents more engineering constraints than behaviourism, as they also include internal processes and, therefore, the nomological boundaries necessary for the MUP would include scientific studies of internal processes as well as external processes. As standard physical properties and processes are multiple realizable and no further physical property needs to be understood, then MUP is metaphysically feasible, but more constrained within engineering (there is more to replicate). See sections 2.5.1.2., 2.5.1.3. and 2.5.1.4.

Standard physical properties that focus on internal processes are now considered. First, those that identify the Ll processes (e.g., neurons firing) with Hl processes (e.g., representations) are discussed (Ll = Hl). Hl-processing has been presented (see section 2.6.2) as at the nomological level of mind/psychology, which include, but need not be limited to, information processing, representations, symbols, computations, and so on. Within the following section of this chapter (section 4.2), a more detailed discussion regarding processes (as opposed to properties) as the necessary and sufficient conditions for minds and persons is given. Within the current section, however, Hl is a category that may include any property or process that is designated mental (mind) from a physicalist perspective. As Ll processes occur in standard physical objects, such as brains, they are

multiple realizable and present less metaphysical constraint as there is only one nomological level to consider.

The current thesis has allocated the category of Ll = Hl as offering less engineering constraints as a clearer understanding of the standard physical processes, and our current technology of replicating these processes and replication of these physical processes (e.g., in neural networks) is advancing within these fields. However, it could be argued that Ll may yet require biotechnological solutions (e.g., neuromorphic computing) and the field of biotechnology is less well advanced (i.e., biotechnology is less advanced than artificial technology such as computers) and, therefore, it could also be argued that Ll = Hl offers more engineering constraints. As stated earlier, the constraints of less and more are placed on a spectrum and the opinion of the current thesis is that artificial artifacts are likely to suffice and, therefore, the allocation to the fewer engineering constraints column is justified. Whether less or more engineering constraints, irrespective of preference for artificial technology or biotechnological (A or B options), is an engineering problem that awaits further empirical research, while offering no metaphysical difficulty if Ll = Hl is adopted. The category of Ll → Hl, including views such as epiphenomenalism and illusionism, indicates that replication of the Ll would result in the Hl and, therefore, also emphasises that replication of the Ll is a necessary and sufficient condition for replication of a mind. Therefore, the metaphysical and engineering constraints for categories Ll → Hl and Ll = Hl are the same, as in both options only the Ll is needed to be understood and replicated (see sections 2.5.2.1, 2.5.2.2, 2.5.2.3, 2.5.3.4, 2.5.2.6, and 2.6)

Next consider the category Ll ←→ Hl, where a weak emergence stance is upheld (the strong emergence stance is dealt with separately below under further physical properties). In this weak emergence stance, the Hl is a standard physical process that occurs at a higher level that, in turn, impacts on the lower levels (similar to how solidity impacts on the motion of atoms at lower levels). In this stance, both nomological levels of psychology (Hl) and neuroscience (Ll) are involved and thus, increase the metaphysical constraints and engineering constraints. In the weak emergence Ll ←→ Hl category, the Hl is deemed to be the level at which causal roles of mental properties are discussed, which are, in some sense, distinct (not metaphysical distinction, but a distinction of complexity of physical processes) from the Ll causal processes, although the Hl nevertheless interacts with the Ll causal processes. For example, a representation may be the product of lower level neurological activity (the Ll → Hl aspect of Ll ←→ Hl), yet these representations interact with each other in a distinct manner giving rise to a representational level of interaction and this representational level of interaction feeds back to the causal processes of the level of neurons firing. An analogy used previously (presented in section 2.6.2), is that between a single bee's

flight and the same bee flying in a swarm. In this analogy, the concept of flight (a purely physical process) occurs at different levels and has resultant processes that can be understood and described at different levels of complexity and interaction (see section 2.6.).

Next, the internal processes category within standard physical substrates is represented by the row, where Hl may be understood as a relatively autonomous nomological boundary. This is not a denial that Ll → Hl, but an acknowledgement that the nomological level of psychology (Hl) provides a sufficient explanation of the human mind and that Ll is not necessary to be understood for artifact production. Because there is only one nomological level to be understood to produce an artifact, there are fewer metaphysical and engineering constraints. This row is analogous to the example of a carpenter who has no need for understanding atomic level interactions that allow for the solidity of wood, even though these atomic interactions are the cause of the solidity. The carpenter can develop theories of different woods, nails, building techniques, and so on that allow for artifact production of chairs, desks and the like. Because there is only one metaphysical nature that needs to be understood (that of standard physical properties and processes at a particular level – here the Hl), this row offers fewer metaphysical constraints.

It could be argued that the psychological level is a long way from being properly understood (e.g., toward a unified consensus theory of psychology), which would indicate that this level is poorly understood (in contrast to the Ll, which can be viewed as better understood at present[46]) and this may indicate that there are some theoretical constraints not yet acknowledged. From an engineering constraint perspective, however, Hl as an autonomous nomological level is deemed to be less constrained, because there is only one level to be understood and replicated. For example, those who are attempting to produce artificial intelligence today largely base this on the idea that minds are computational systems (however defined) and, therefore, that this is the necessary and sufficient condition irrespective of whether that computation is performed by biology (as it is assumed the human mind is) or that of a silicon-based substrate, such as a computer.

The following two rows under physical substrates with standard physical properties with an internal process emphasis represent the categories of neurocentric and embodied cognition. In both of these rows, the mind is deemed to be a result of standard physical properties and their processes and, therefore, both offer less metaphysical constraints. The distinction between these two arise in relation to engineering fewer constraints, because the neurocentric view asserts that

46 Whether the Ll is better understood could equally be questioned but it appears, at present, that there is at least a standard epistemology (the scientific method) and agreed upon properties (e.g., energy, mass, and so on) and, therefore, the current thesis views the Ll as a better understood level.

only part of the substrate is necessary and sufficient for the mind, whereas the embodied cognitivist expands the necessary and sufficient conditions to the entire (or at the very least more of the) substrate. Because the embodied cognition view requires more of the substrate to be replicated, this would increase the engineering difficulty (see section 2.5.2.5)

The second main category under physical substrates is that of further physical properties, which, in turn, is further categorised into the categories of strong emergence and panpsychisism. Both the strong emergence and panpsychic views assert that the mind is the result of physical properties and processes and, therefore, it is likely that the mind is multiple realizable (see the multiple realizable argument in chapter 2). Because both of these views assert that there are further physical properties, this increases the metaphysical ground that needs to be covered to understand the mind and, therefore, offers more metaphysical constraint for the MUP. As all further physical rows share the column of more metaphysically constrained, the differences arise in relation to engineering constraints. Strong emergence is further categorised as either Ll ← → Hl or as Hl (autonomous). Within the Ll ← → Hl row, there is more engineering constraint, as the MUP would need to consider both the L and the Hl properties and processes since both nomological levels (the Ll and the Hl) would need to be included in the artifact construction. Within the Hl (autonomous) row, there are fewer engineering constraints, as understanding the Hl is sufficient for replication. For example, in relation to artificial intelligence, it could be argued that computations result in strong emergent properties (see below) but as long as both properties and processes are the same (e.g., the same strong emergent representational properties interact in the same manner), and both emerge from Ll physical processes (e.g., the Ll of neurons in biology and the Ll of silicon based hardware in computers), then the same mind could, in principle, be replicated in terms of multiple realizable physicalism. By asserting that the Hl contain further physical properties, there is greater metaphysical constraint (there is an ontological assertion of some new physical property), but because how this is achieved is not essential to replication (only the Hl is needed for replication), there are fewer engineering constraints (see section 2.7).

In relation to panpsychism (which asserts that mental properties may occur at all levels), there is a greater engineering constraint, in that these properties are little understood and how an artifact could include these properties appropriately is also less understood. Chalmers (1996, 2014) presented the dancing qualia argument (see section 2.7.1) as a means to establish some form of method to explore how the physical functions (standard physical properties and processes) could retain consciousness across substrates. In short, parts of the brain are replaced gradually and consciousness is either retained, lost, or dances between conscious and non-conscious states.

The final two rows represent the options of multiple realizability and non-multiple realizability. All physical properties and processes are allocated by this thesis's multiple realizable physicalism as multiple realizable (based on the multiple realizable arguments from chapter 2) and non-physical substrates as non-multiple realizable. Thus, in essence, Table 4-1 presents an overview of chapter 2 and reaffirms that if a substrate is physical (however defined) it is multiple realizable and replicable, allowing for the feasibility of the MUP.

Before progressing to the table of chapter 3 (Table 4-2), a brief comment is made here on some of the complexities of philosophy of mind debates and the table presented in Table 4-1. The aim here has been to simplify the field in relation to the MUP. The current thesis acknowledges that certain philosophies of mind may include an intersection of the options presented. For example, functionalism, as it applies to philosophy of mind, traverses some of these distinctions. As was presented in chapter 2, there are higher level functionalist views (e.g., second order functionalism, role-functionalism, functional states identity theory, and so on) and lower level functionalist views (e.g., first order functionalism, realizer functionalism, functional specific claim). The lower level functionalist views align with the Ll = Hl rows, whereas the higher-level functionalist views align with Hl (autonomous) and Ll ← → Hl rows. Furthermore, functionalist views can also include external processes, such as behaviours (in chapter 2, this was presented as "environment ← → [Ll ← → Hl]" or some variation thereof – see section 2.6 and Figure 2-3). In these instances, functionalism does not directly correlate to a particular row and, thus, a particular philosophy of mind may be applicable to multiple rows, depending on the particular preferences of the philosopher.

Alternatively, other philosophies, such as type identity theory, can be readily aligned with the Ll = Hl row (the mind is identified as the biological process such as neuron interaction). What the table hopes to achieve is, therefore, not a complete correlation of all philosophies of mind (here focusing on the mind–body problem), but rather a heuristic of key emphases of a complex field as it relates to the MUP. As the MUP is focused on replication/uploading minds through artifact creation, the table presents an overview of the options available to the MUP and allows for combining aspects as the philosopher sees fit. What has been demonstrated here and throughout chapter 2, is that multiple realizable physicalism (as defined in the current thesis) serves as unifying concept that may align with multiple variations of solutions to the mind–body problem that allows for the feasibility of the MUP. This section now turns to chapter 3 and presents a similar table with the similar aim of sketching a heuristic of the persistence problem as it relates to the MUP.

The mind–body problem									
Distinctions				**The MUP is ...**		**Constraints**			
				Non-feasible	**Feasible**	**Metaphysic constraints**		**Engineering constraints**	
						Less	**More**	**Less**	**More**
Non-physical substrate				✓					
Physical substrate	Standard physical properties	External processes	Behaviourism		✓	✓		✓	
			Externalism/extended mind		✓	✓			✓
		Internal processes	Ll = Hl		✓	✓		✓	
			Ll→ Hl		✓	✓		✓	
			Ll ←→ Hl (weak emergence)		✓		✓		✓
			Hl (autonomous nomological boundary)		✓	✓		✓	
			Neurocentric		✓	✓		✓	

			Embodied cognition		✓	✓			✓
	Further physical properties	Strong emergence	Ll ←→ Hl		✓		✓		✓
			Hl (autonomous)		✓		✓	✓	
		Panpsychic			✓		✓		✓
Multiple realizable					✓				
Non-multiple realizable				✓					

Table 4-1: The mind–body problem and the MUP

4.1.2 The psychological solution and the feasibility of the MUP

Table 4-2 presents a summary of chapter 3. The main thrust of this chapter was to show that the biological solution to the persistence problem leads to the MUP being non-feasible and the psychological solution as being feasible. The reader is reminded that the current thesis has opted to focus on intrinsic properties or processes playing a determining role in personal identity and its persistence through time, as opposed to arbitrary/extrinsic properties or processes. Although the current thesis is convinced that persons are intrinsic processes, the arbitrary/extrinsic option offers no difficulty for the MUP (e.g., if continuity of persons are assigned through extrinsic means then the MUP would simply need to occur in an appropriate community such as a trans-humanist one that would likely make the extrinsic attribution of personal continuity).

With the focus on intrinsic properties and processes, the reader is further reminded that, as per previous arguments presented in chapter 3, neither the biological nor the psychological solution claim absolute identity as necessary for the persistence of persons and that both refer to some form of partial identity. Once partial identity is accepted, the question arises as to what features/conditions/qualities/properties/ processes are to be deemed necessary and sufficient for the persistence of persons. The biological solution emphasises that space-time continuity of the substrate (the body) is a necessary and sufficient condition, whereas the psychological solution emphasises that it is the continuity of psychology that is the necessary and sufficient condition.

The psychological solution as presented in the literature (see chapter 3) is often presented as the psychology continuing across different substrates. For example, in Locke's (1689) prince and cobbler thought experiment, the psychology of the one person continues in the body of the other. In Parfit's (1984) teleportation, the person continues even if the psychology continues in a new substrate on a different planet (Mars in his thought experiment). The psychology solution, therefore, presents with an alliance with the MUP, in that both assert the feasibility of persons continuing in an alternative substrate, so long as there is psychological continuity. Therefore, the psychological solution, when applied to the MUP, is presented as the necessary and sufficient metaphysical condition for the MUP. This section now describes how Table 4-2 is presented and how rows and columns are to be understood.

First, consider the rows of Table 4-2, indicating the categories of the "process/psychological solution" and the "substrate/biological solution". As with Table 4-1 that dealt with the mind–body problem, Table 4-2, in its turn addressing the persistence problem, emphasises options that present with less or more metaphysical constraints. Within Table 4-2, the metaphysical constraints

are presented in rows and as further categories of the psychological solution. The biological solution is not deemed less or more metaphysically constrained, as the solution to the persistence problem negates the metaphysical feasibility of the MUP and, therefore, constraints of less or more feasibility are not appropriate. As with the previous table, in this section there is also a designation for engineering but, in Table 4-2, these are allocated to upload scenarios rows titled "destructive" and "non-destructive". These upload scenario rows were deemed appropriate, because the various columns (to be discussed below) impact on which upload scenario is to be preferred, depending on the emphasis of each column.

Second, consider the columns of Table 4-2 where two primary columns are placed, namely "There can be only one? Replication" and that "The MUP is". "There can be only one? Replication" column is further categorised into the various alternatives discussed in chapter 3, namely; the numeric/qualitative identity views, the token–type distinction, and the branching/non-branching distinction. As will be discussed, not all of these distinctions deny the feasibility of the MUP, which is the second primary column (further categorised as the MUP is feasible/non-feasible). Furthermore, the distinctions within "There can be only one? Replication" do have an impact on the preferred upload scenario. The distinctions within "There can be only one? Replication" is the core of the current section and, therefore, whereas in Table 4-1 each row was discussed sequentially, in Table 4-2, each column is discussed sequentially. As with Table 4-1, Table 4-2 does not aim to establish which of the options is veridical but to present a map as to how options within the persistence problem present in relation to the MUP.

Within the "There can be only one? Replication" column, first consider the numeric/qualitative identity distinction. Within the numeric identity claim category, there can only be one numeric person and, therefore, the question arises as to what numeric condition the identity of a person should be attributed to. As the two primary options within the persistence problem is to either identify the person with the body or the psychology these options intersect within the numeric distinction. In the numeric psychological solution, the MUP is metaphysically feasible because continuity of psychology (irrespective of substrate) is asserted, however, there can only be one (numeric identity) continuity at a time. This aligns the numeric psychological continuity stance with destructive uploads that allow for only one continuity of psychology to occur at a time. This can be done gradually (e.g., where parts of the brain are gradually replaced) or suddenly (e.g., in the scan and copy scenarios), but results in the same outcome of an artifact substrate that instantiates a person who was originally instantiated in a traditional biological substrate (the body). See section 3.5.

This raises philosophical difficulties for the adherent of this view, because there is now a metaphysical claim (numeric identity of persons) that is dependent on what seems to be an arbitrary event (whether one or two replicas are made) as well as other difficulties, such as the fission problem (presented in section 3.4). These philosophical difficulties are the reason the current thesis allocated the numeric psychological view as being more metaphysically constrained. In this option, the MUP is, nevertheless, feasible and so allocated in the relevant column.

The biological solution is inherently a numeric identity claim because it asserts that the particular biological substrate (numerically allocated as a space-time continuity of substrate) is the necessary and sufficient condition for continuity of persons. In this view, the numeric identity is attributed to the numeric body and the feasibility of the MUP is denied. The current thesis in chapter 3 argued that numeric identity is primarily a confusion between substrate space-time continuity and space-time instantiations of processes (see section 3.5). Furthermore, the thesis has argued that in relation to persons, qualitative identity is the appropriate approach. When qualitative identity intersects with the psychological solution (as is the preferred option of the current thesis), this allows for both destructive and non-destructive uploads (whether the original is destroyed, or whether multiple replicas are made is irrelevant because, so long as there is the continuity of psychological qualities, there is a continuity of the person). This option is deemed less metaphysically constrained as the qualitative psychological solution encounters fewer problems (e.g., it is not arbitrary, it solves the fission problem, and so on) than the numeric psychological option as well as aligning to a greater extent with previous chapters' view on minds as multiple realizable (i.e., if minds are multiple realizable and the psychological solution is upheld why should persons not be multiple realizable?).

This is not to say that the qualitative psychological solution is not without its own problems but that it is the view of the current thesis (as argued in chapter 3) that these can be minimised. In this option, the MUP is, therefore, feasible and so allocated in the relevant column. In contrast to the possibility of a qualitative psychological solution, the biological solution cannot accept the qualitative identity, as the numeric substrate is, in this view, the essential condition for continuity of persons. In the biological view, qualities may be realized across substrates (e.g., pain can occur across substrates), but persons cannot continue across substrates and numeric identity is the preferred and only legitimate stance in relation to persons. Therefore, this view implies unfeasibility in relation to the MUP.

Next, consider the type–token distinction. In chapter 3 and, therefore, in this section, the token refers to a particular substrate whole, where the type refers to the particular person[47]. As discussed in chapter 3, the token can either be identical with the type (type = token), distinguished from the type (type ≠ token), or the type can be denied. Denial of the type or accepting the type = token amounts to the same outcome, where only the token need be affirmed. When presented this way, if a type = token there can be no instantiation of the same type (here a particular person) across alternative substrates as the token and the person are one and the same. This view is, therefore, synonymous with the biological solution as the particular physical token (the body) is taken to be precisely the same as the type and they cannot be distinguished. In Table 4-2, the type = token is, therefore, allocated to the biological solution row only and no engineering options are allocated as this is not an MUP feasible option.

In contrast, if it is asserted that type≠token, then the type may be instantiated across multiple tokens. If a person is a psychological type and this type were to be instantiated in an alternative token (i.e., an alternative substrate), then the MUP would be feasible. Furthermore, as the type can be instantiated in an alternative token then it could in principle be instantiated in two alternative tokens. Therefore, the type could in principle be instantiated across multiple tokens and, therefore, would allow for both the destructive and non-destructive upload scenarios. There is, therefore, a natural alliance with qualitative identity claims and type≠token claims, in that they both allow for certain conditions (here psychological conditions) to be instantiated across substrates and, therefore, allow for both destructive and non-destructive upload scenarios (see section 3.6).

Lastly, the distinction between branching and non-branching is considered. Similar to the qualitative and type≠token options, the branching option allows for multiple replicas, whereas the non-branching allows for only one continuity (such as in numeric identity). The same difficulties arise for those who assert that persons are non-branching as with the numeric option, in that it appears to offer an arbitrary assertion that there can only be one person in a particular substrate (as well as having difficulty with problems, such as fission). For those who retain both the psychological solution as well as the non-branching option (the non-branching psychological option), the MUP is feasible, as the person may continue in an alternative substrate, but this category is more

47 The reader is reminded of the type sortal spectrum introduced in chapter 1 and further developed in relation to persons in chapter 3. The type–token distinction within the persistence problem refers to the type of a particular person and whether this person can be instantiated in different tokens. This is distinct from the type–token distinction as it relates to the mind–body problem, where variation of neural tokens within the same substrate (substrate parts) can be understood as types.

metaphysically constrained due to the problems mentioned above (e.g., fission). However, as there cannot be any branching within the non-branching psychological option, the MUP must pursue destructive upload scenarios.

The non-branching option can also be applied to the biological solution, in that there can only be one person (who is identified as the numeric substrate of the body) that continues in space-time. However, the biological non-branching option differs from the non-branching psychological option because the biological non-branching option denies instantiation of the person across substrates (the person is the substrate).

Now consider the branching psychological option that allows multiple branching of the same ancestral person. This option is less metaphysically constrained as it accounts for some of the problems within philosophy (e.g., fission) as well as being congruent with a more meaningful account of what qualities are essential for the continuity of persons, such as space-time instantiations, dynamic-open, control, and self-indexing systems (which, in turn, could account for the intuitions that support the non-branching and numeric claims – see section 3.5). Furthermore, the branching psychological option allows for both destructive and non-destructive uploads, as the continuity of persons is not dependent on the outcome of only one psychology continuing (i.e., the essential condition is not the number of continuing instantiations but simply that that there is psychological continuity). See section 3.7.

Because all psychological options (irrespective of whether one holds to branching or non-branching, numeric or qualitative options) allow for the MUP, the psychological solution to the persistence problem is the necessary and sufficient condition for the MUP. Furthermore, perspectives that offer fewer metaphysical and engineering constraints to the MUP are qualitative identity, type ≠ token, and branching options. These less constrained options have been argued as veridical by the current thesis, although it is acknowledged that other options (e.g., numeric psychological, non-branching psychological) may also be applicable to the MUP.

This section has presented two tables that indicate that the MUP is feasible within multiple philosophical stances and that the two primary necessary and sufficient conditions are that of multiple realizable physicalism (in relation to the mind–body problem) and the psychological solution (to the persistence of identity problem). The thesis now turns to a novel way to integrate these problems within the context of the MUP, namely the thesis's emphasis on the mind in terms of process.

The persistence problem and the MUP												
			There can be only one? Replication						**The MUP is**	
			Numeric	**Qualita-tive**	**Type = token (or denial of type)**	**Type ≠ token**	**Branch-ing**	**Non-branch-ing**	**Non-feasible**	**Feasible**
Primary emphasis	Process/ Psychological solution	Less metaphysical constraints		✓		✓	✓			✓
		More metaphysi-cal constraints	✓					✓		✓
	Substrate/ Biological solution		✓		✓			✓	✓	
Engineering (upload scenarios)	Destructive		✓	✓		✓	✓	✓		
	Non-destructive			✓		✓	✓			

Table 4-2: The persistence problem and the MUP

4.2 The process-self

The current thesis is concerned with the feasibility of the MUP and, therefore, certain areas of enquiry have been emphasised, namely, what is the nature of minds and persons such that they could, in principle, be instantiated in an alternative substrate. A primary distinction of this thesis has been between the substrate and what the substrate does (processes). This idea has been associated with Sellars' (1963) objective of philosophy as to explore things/entities and how things hang together/processes with, in the sense of the thesis, the substrate being the 'thing' (including the various parts of the substrate which are also things) and 'how things hang together' being the processes that are instantiated by/in the substrate. A thing may be inactive, whereas a process is necessarily active.

This distinction emerged in both chapters 2 and 3 under the corpse problem, where a body may be seen to exist without a mind/person. Importantly, the thesis has taken this to reflect the substrate (in this case a body) as a thing and the processes performed by said substrate being identified with the mind/person. Throughout the thesis, the category mistake of categorising minds and persons as entities (e.g. substrates such as bodies) when they should be processes (e.g., what the substrate does), has been presented in multiple stances across both the mind–body problem and the persistence problem. In relation to the mind, the behaviourist states that minds are the external processes and the identity theorist views minds as the internal processes (e.g., C-fibre stimulation and not C-fibres), while the functionalist refers to the causal processes as the mind, and so on. In relation to personal identity, the psychological solution implies that persons are minds and that introducing or inserting a further category between minds and persons is another category mistake (e.g., the assertion that persons are bodies rather than minds, or at the least a combination of minds and bodies).

The current thesis has argued that minds are processes, and persons are these minds (the psychological solution), and then that persons and minds are the same processes. There is the possibility of arguing that mind processes and person processes are distinct processes and, therefore, there may be a need to establish that they are in essence the same processes. This possibility has not been explored within the current thesis, as the two solutions of continuity of persons (biological and psychological) do not emphasise this possibility.

The process solution suggested in this thesis can thus be presented as:

(1) Mind = collective processes A (multiple realizable physicalism)

(2) Person = collective processes B (the psychological solution)

(3) Collective processes A = collective processes B

(4) Therefore, mind = person

If minds are processes (as asserted by chapter 2) and persons are processes (as asserted by chapter 3), it is likely that these are the same processes. This assertion could, in principle, be rejected, but because there is no other process generally emphasised within personal identity (i.e., the two primary options are of focusing on the body or the psychology), these are taken by the current thesis to refer to the same processes. In essence, if one rejects the biological solution to personal identity, as well as non-physical entities (such as the soul), it is unclear what process is left for persons to be other than psychology (the mind). And therefore, it is implied by those who hold to the psychological solution that the processes attributed to persons are the same processes as that which are attributed to minds (minds and psychology being synonymous terms in this context).

The thesis here uses the term 'self' to act as a category for both a particular mind and a particular person. The continuity of the process-self (both mind and person) is, therefore, used to present the collective processes that constitute a mind and the collective processes that constitute a person over time. The basis of the concept of the process-self, is that what matters to minds and persons are the processes and not the substrate that instantiates these processes. The processes that are presented in this section are those that occur within a physical substrate and its environment[48] and, therefore, physical processes are at issue. Furthermore, the aim of this section is not to explore which nomological boundary of processes are necessary and sufficient (e.g., external processes, internal processes, Hl, Ll, or any combination of these) for personal identity, but rather to assert that physical processes are the appropriate identity category for the self. Therefore, the current thesis views the self as the physical processes of the physical substrate, and that it is these physical processes that are to be identified as the self rather than the substrate (e.g., the body) that performs these processes.

As discussed in previous chapters, physical processes require physical substrates and are instantiated in space-time. For example, a wave function (a process) cannot occur without a physical substrate to be instantiated in (e.g., the physical properties of the slinky are needed to instantiate the wave process, whereas a brick could not instantiate the wave process), however, there is no such 'thing' as an inactive wave function if the substrate is not in an active state (e.g., if the slinky

48 It would also be possible to affirm a process-self from a substance dualist perspective. For example, consider a soul made of some ethereal spirit molecules substance. This substance is the substrate of the mind/person, but in the process-self variation of substance dualism, a self would only exist if this soul substrate instantiates a process.

were to be inactive). Simply put, there can be no physical process without an appropriate physical substrate, but the identity claim for the self is that it is the processes that matter and not the substrate that instantiates the processes.

Within the current thesis a distinction between properties and processes has been made to allow each philosophical stance to use terms from their own perspective. It has been demonstrated that whether one holds to standard physical properties (e.g., identity theorists) or further physical properties (e.g., property dualists), that these are likely to be replicable and, therefore, allow for the feasibility of the MUP (multiple realizable physicalism). The process-self makes a further claim that, although physical properties (however defined) are necessary to produce processes, it is the processes that are the appropriate metaphysic identity claim for selves (minds and persons). For the MUP, which emphasises the distinction between substrates and processes, the types of properties that are emphasised are those that can occur without a process at the given nomological level under discussion. Properties, therefore, describe what the substrate is without any activity at the desired nomological level and the processes in terms of what the substrate does at that nomological level. The property defines the thing and the processes defines what the thing does.

The current thesis acknowledges that the term 'property' can have multiple applications and, therefore, the use of the term in this section requires some clarification in relation to the MUP and the process-self. A property may be defined as "a quality, attribute, or characteristic that belongs to something" (Mautner, 2005 p. 500). This idea of belonging/having can be applied to both entities and processes. For example, I *have* a finger (i.e., the body may be said to be constituted by many discernible things, such as can be described in any anatomy text book). If all of these things are removed, the body ceases to exist. In this sense, a property can be seen as an entity; a thing (or description of a thing) that then accumulates to form another thing. Now consider that a property can also be considered as a process that belongs to you. For example, I can *have* team spirit (e.g., as used by Ryle to explain his behaviourism) or *have* pain (e.g., as described by identity theorist in relation to C-fibre stimulation). In this sense, pain and team spirit are attributes, qualities, or characteristics (i.e. properties) that belong to me (and, therefore, can be described as properties), yet they may also be described as processes (here assuming that there can be no team spirit without a process within a team, or pain without a collection of biological processes).

Properties have also been described as processes elsewhere. For example, Orilia, and Swoyer (2020) contrast properties with relations (a form of process – see also MacBride, 2016), but then acknowledge that as both may be predicables and exemplifiables, that relations "may even be viewed as kinds of properties". Furthermore, consider the distinction between extrinsic and intrinsic properties (Marshall and Weatherson, 2018), where an extrinsic property can be attributed

to the way an entity interacts with the environment. For example, if I am married, then this is not deemed an intrinsic property to being me (if I cease to be married I do not cease to be me) as it is the external circumstances of an ongoing relationship to another thing (e.g., a wife) that maintains the 'property" of being married. Within the scientific literature, properties may also refer to relations. For example, in chemistry, a physical property can "describe the physical characteristics of a substance" and a chemical property can "describe how a substance reacts with other substances" (Moore, Hren and Mikulecky 2015, p. 211). In this chemistry perspective, physical properties are related to the substrate, whereas chemical properties may relate to processes within or between substrates. Therefore, the term 'properties' may be attributed to processes as well as substrates, in which case, the property-process distinction made by the current thesis would be less meaningful. It is not the agenda of the current section to resolve many of the ongoing philosophical disputes in relation to properties (see e.g., MacBride 2016, Marshall and Weatherson, 2018), but to enquire into the metaphysics for the MUP and, therefore, below presents one stance that upholds the property-process distinction and may be useful to the task at hand.

As stated earlier in this section, the MUP is concerned with substrates and their processes and the idea that an alternative substrate may continue a process. Consider the analogy of a standard computer, where the hardware is a substrate and software are processes. The hardware substrate has properties of mass, weight, chemical compounds, and so on, which are discernible whether the computer is active or not (e.g., it is not plugged in). Properties, at this nomological level[49], exist without any process. Now consider the software that is active when a particular program is running on the hardware. The software, so viewed, is an active process. If the computer is turned off, the software ceases to exist (although its potential to exist may be stored), as the ontological claim is that the software is the activity of information processes. The appropriate identity claim for software is, therefore, attributed to the process and not the substrate hardware. That the software is a process instantiated in hardware does not negate that need for the hardware to have certain properties, such as particular circuitry architecture (e.g., a standard toaster will not run a software program) and that it is space-time instantiated (e.g., if there is no appropriate hardware in space-time no software process could exist). Properties are, therefore, necessary to instantiate the processes but the identity claim of what software is, is that software is not these properties but the appropriate processes.

---

49 Within any given standard computer there are many active processes occurring at other levels (e.g., quantum levels) but these are deemed not relevant to the artifact creation and maintenance of standard computers and software programs and so this nomological boundary may be asserted.

The nomological boundary may act as a means to define what level of substrate and substrate interactions are needed to understand (and replicate) a phenomenon. For example, at the atomic level, the substrate things are electrons, protons, and so on that interact, but for the carpenter, the substrates are wood, nails, hinges, and so on that interact. The current thesis has acknowledged that in relation to the mind, there may be various nomological levels put forward (e.g., the substrate whole with its environment, the internal substrate processes such as neurons firing, and so on) and does not attempt to resolve which level is needed. What has been demonstrated, is that so long as the substrate is physical and the processes are physical, then replication (and by inference the MUP) is feasible. The concept of the process-self then has made the further assertion that what matters to identity (what is a mind, what is a person) is that the processes, at whatever nomological level is deemed appropriate, are what is to be identified as the self.

In relation to the mind–body problem, the property-process distinction can then be seen as properties relating to the substrate of the body and processes as relating to the mind. In the persistence problem, properties can be associated with the body (the biological solution) and the processes as relating to the psychology (the psychological solution). Properties, understood in this way, are attributed to substrates and amount to a negative hypothesis, where properties are that which can be attributed only when there are no processes at the appropriate nomological level (within both the mind–body problem and the persistence problem these relate to the biological substrate of bodies and the psychological/mind processes).

Properties in this view are inactive, or passive, in that a substrate may possess properties while being inactive and properties do not initiate processes but allow for their instantiation. A body substrate may have multiple inactive properties such as mass, anatomical structure, and so on. All of these properties exist without the substrate being active (e.g., when the substrate is a corpse), yet it is deemed that these properties would be necessary for the processes to occur when a living person is present. This view on properties is similar to Kim's (2010, p. 301) view, where "... intrinsic properties and, hence, causal potentials ..." indicate that properties are the precursors to processes. Armed with this clarification on the linguistic convention of the property-process distinction as it relates to the MUP, the thesis now turns to the concept of a process-self as a unifying concept of the mind–body problem and the persistence problem illustrated in Figure 4-1.

The process-self is a broad concept because it does not predetermine which physicalist stance (e.g., which nomological boundary is necessary and sufficient) is preferred so long as it is the physical processes (instantiated by physical substrate and the environment) that matter for the self. The current section does not present a detailed argument as to why a process-self should be

preferred but aims to illustrate how a process-self view can integrate the mind–body problem and persistence problems in relation to the MUP.

Figure 4-1 places the mind (as it relates to the mind–body problem) on the left of the diagram and the person (as it relates to the persistence problem) on the right of the diagram. Three rows are superimposed over the flow diagram with the categories presented from bottom to top of "The mind is not replicable", "The substrate is replicable. The mind is replicable. The person is not replicable", and "The mind is replicable. The person is replicable".

In relation to substrates, Figure 4-1 presents two options for the substrate as either being constituted by physical or non-physical substrates. A non-physical substrate (as discussed in chapter 2 under substance dualism) would indicate that the self falls beyond the purview of any science and, therefore, would not be replicable by any artifact. The current thesis (see chapter 3) in relation to persons, given the two primary distinctions within the literature of the biological and psychological solutions to personal identity, did not address the possibility of a non-physical person (i.e., a soul or spirit of sorts) and the person as being instantiated or identified with a non-physical substrate and, therefore, has no allocation within the diagram. Within the category of a physical substrate, two types of properties are considered. First, the standard physical properties (e.g., weight, length, chemical composition) and then the further physical properties (e.g., phenomenal properties, qualia) are considered. As both these types of physical properties are deemed to be replicable, due to multiple realizable physicalism in terms of the argument made in section 2.2, both allow for the replication of the substrate. If a substrate is replicable, then it can be inferred that the potential processes (what the substrate could do) would also be replicable (the supervenience principle), as illustrated in the top row of the diagram.

In standard physical processes, the process-self may include external processes, such as in behaviourism. Behaviours are, by definition, processes (there can be no inactive behaviour) and for the MUP, a substrate that may instantiate these behaviours would be said to be a continuity of that self. Second, consider internal processes that emphasise the Ll processes (e.g., pain being C-fibre stimulation and not C-fibres), then these processes can, in principle, be replicated in an alternative substrate. Here, there are internal properties that are needed for the processes to occur (e.g., the properties of C-fibres), but pain is only instantiated when there is firing (processes). This could include views such as epiphenomenalism and illusionism, where the phenomenal first person subject experiences is non-causal but are, nevertheless, is the result of Ll processes. For those who opt for Hl processes (e.g., Hl functionalism), again there is no self unless there are processes occurring (now drawing the nomological boundary at the Hl) at the higher level.

Further physical property substrates assert the ontological metaphysics of phenomenal properties (e.g., WIL, qualia, raw feelings). These further physical properties are deemed by those who assert them (e.g., property dualists) as having the feature of being intrinsic, which in the context of these views, means non-relational properties. Therefore, there is a distinction in the further physical property view between properties (here phenomenal properties) and processes (relations). The phenomenal properties may be viewed as the building blocks from which a mind is constructed, the mental atoms as if it were. Therefore, those who hold to phenomenal properties, assert the ontological reality of these properties, but need to further assert that there are nevertheless processes (relations) that also occur for a mind to be present.

The process-self view asserts that it is these processes that make up the identity claim for the self. In the emergent view, emergent phenomenal properties are the result of Ll processes and, therefore, the replication of the Ll will lead to these Hl phenomenal properties (supervenience principle). However, these Hl properties do not constitute a self unless there is also processing occurring at the Hl and it is these processes (the interaction between Hl properties) that are deemed by the process-self as to be the metaphysical nature of selves.

The same applies to panpsychists in the process-view, in that even though here there is an assertion of some further physical properties (i.e., phenomenal properties), in themselves, these do not constitute a mind unless there is a process instantiated. This may account for why property dualists (e.g. Chalmers, 1996, 2014) include functions (processes) as part of their philosophy of mind. Furthermore, as panpsychism asserts that these phenomenal properties are present in all (pan) things, then they should be present in alternative substrates. It is acknowledged by the current thesis that these are complex issues and require further clarification and exploration. However, because the current section merely aims to present the concept of the process-self, rather than exhaustively defend these claims, this falls beyond the scope of the current thesis and awaits further work in this regard at a later time. The process-self view, therefore, asserts that properties (although necessary) are insufficient for establishing minds and that it is the processes that are deemed the appropriate metaphysics of minds and persons.

In considering the persistence of persons (the right side of the Figure 4-1), the distinction of the biological is associated with the substrate and the psychological is associated with the processes. The current thesis (in chapter 3) offered three options in relation to substrates and processes of, 1) substrate only, 2) both substrate and processes, and 3) processes only, and these are reflected in the diagram. In this section, as per the previous discussion, properties are aligned with substrates (what the entity has) in contrast to processes (what the entity does).

Within the biological solution, it is the substrate that matters for personal continuity with the substrate-only option (1) as relating to the substrate without any processes. In this stance, the substrate may be replicable but the person is not. Option 2), 'both substrate and processes', requires some explanation as processes are mentioned, but elsewhere in the diagram (the top row of the diagram) there are internal and external processes. The diagram is not asserting that the biological solution to the persistence problem is positing a different kind of process than that of internal and external processes at the top row of the diagram. Rather option 2),'both substrate and processes', indicates the biological solution's insistence on the substrate as being the primary focus for identification of persons even if processes are included. In option 2), 'substrate and processes', of the biological solution, the processes are secondary to person continuity, which is primarily identified with the substrate.

Furthermore, in this option 2), the mind (the processes) could be replicated, as demonstrated in the various thought experiments, where minds are replicated in an alternative substrate, but the person is not (see section 3.4.). Therefore, this biological solution (option 2) allows for the replication of a mind but not of persons. For example, a molecule for molecule substrate would (usually based on supervenience) have the same mind. But even when the same processes are active (the same mind is instantiated), the biological solution denies that it is the same person (this is the fundamental difference between the psychological solution and the biological solution as presented in chapter 3). This row, therefore, in allowing the biological solution to assert their own views, illustrates that the biological solution, even when accepting that minds may be replicated, continues to assert that persons are not. The current thesis has taken this to be a category mistake of confusing substrates for processes (see chapter 3) and follows the psychological solution in asserting that persons are minds.

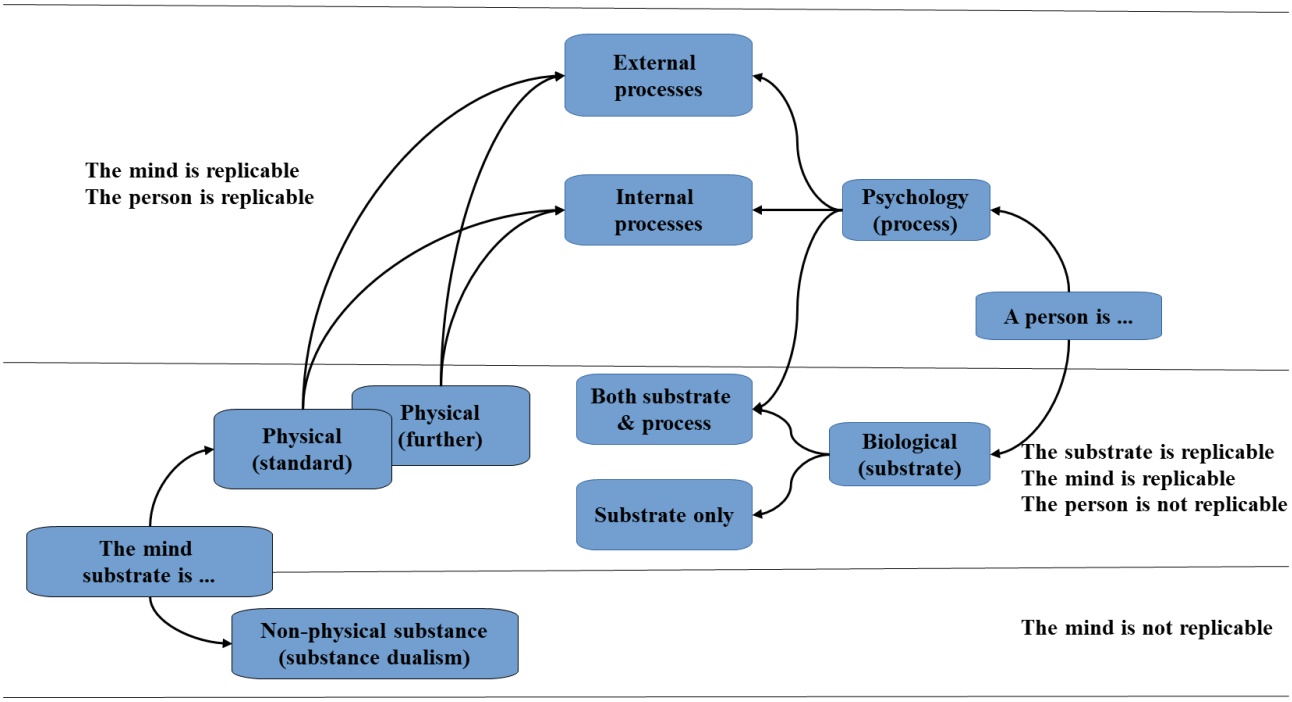

Figure 4-1: The process-self integration

Remaining on distinctions from the right of the diagram (a person is …), in contrast to the biological solution, is the psychological solution which is option 3) of 'process only' in relation to the persistence problem in the table. In terms of the continuity of persons, the current thesis has not explored whether internal or external processes are necessary and sufficient for a person. It sufficed to say that if minds are persons (the psychological solution) and these are both the same processes (in the diagram internal and external processes), then whatever holds true for minds (in the mind–body problem) would hold true for persons (in the psychological solution). If processes are replicable (multiple realizable) and also represent the appropriate metaphysical stance in relation to both minds and persons, then both the mind and the person are replicable.

The notion of the process-self is a broader concept than many of the previously discussed (see chapter 2 and 3) philosophical stances. The process-self allows for the mind to be both causal (such as the DBA model) as well as non-causal (e.g., eliminativism, illusionism, epiphenomenalism). Within the non-causal mind view, the mind is an epiphenomenon of sorts and is the result of Ll causal processes. The Ll processes are still physical processes and, therefore, part of the process-self. Now consider those stances that state the mind is causal. For example, in functionalism within philosophy of mind, the assertion is that the functions (processes) of the Hl are causal (e.g., Hl as an autonomous, nomological, boundaried process, or Ll ⟷ Hl as Hl being causal and interacting with the Ll to produce the mind, or any such variation). The process-self can then be included in

the functionalist view as the functionalist view asserts that it is the functions (here processes) that matter. Therefore, both causal and non-causal views of the mind may be included in the process-self view so long as it is the processes (at whatever level) that are identified with the self.

The concept of the process-self, therefore, includes processes at any designated level (internal, external, Hl and Ll, or any combination thereof), the mind may be causal or non-causal, and may include both standard physical substrates or further physical substrates. As well as offering a meaningful distinction (substrates/substrate processes) between the psychological and biological solutions to the persistence problem, the concept (process-selves) offers a coherent perspective on both the mind–body problem and the persistent problem and integrates multiple philosophical stances. The concept of the process-self, as presented here, offers, therefore, a way to integrate many of the disparate aspects of the mind–body problem and the persistence problem. If minds are viewed as processes (what the substrate does) and are not the substrate's properties (what a substrate is), then it would allow for the MUP. Any substrate that can instantiate the same processes is deemed, in this view, to be the same self.

Although there may be many engineering difficulties (e.g., precisely what are the properties of the substrate that are needed to instantiate the process), the metaphysical problem of the MUP, whether a person can be instantiated in an alternative substrate, is resolved. Furthermore, problems that occur in the mind–body problem and the persistence problem may also be resolved, as has been presented throughout the thesis. For example, the mind–body problem often makes distinctions between the mind and the body, which leads to substance dualists claims that, because minds and bodies are so disparate, that they must be different substances. The process-self view resolves this by affirming that they are indeed disparate but that the differences lies in the differences between substrates (entities) and what the substrate does (processes) and that to categorise a process as an entity is the primary category mistake identified throughout chapter 2. Similarly, chapter 3 presented arguments (e.g., the corpse argument), where the mistake of categorising persons as their substrates, as opposed to the processes of these substrates, is a further category mistake.

The process-self view also offers an explanation as to why the self (mind and person) appears different to our bodies yet is always associated with them. That the self is a physical process means that there needs to be an appropriate substrate for the process to be instantiated in. Because human bodies (at present) are the only known substrates that can instantiate these processes there is a natural intuition to align the person with these processes. The alignment/correlation between mind and body is affirmed by the process-self view (there can be no physical process without a physical substrate), but the identity claim is denied (i.e., the self is a process and not the substrate

that instantiates that process). Furthermore, the complexity of processes would make replication of an appropriate artifact to instantiate these processes extremely difficult and, therefore, this complexity may account for why persons are deemed to be unique (as well as the intuition of numeric identity). However, if what has made persons 'unique', up until this point in history, is the complexity of the process (as in the process-self), then this is an engineering rather than metaphysical problem. Therefore, the process-self view gives an account for why we have the intuition to associate the self with a particular substrate (our bodies), while not falling into the category mistake and resulting identity claim that the self is that substrate. Consider a simplified process-self argument:

(1) Selves are correlated/instantiated in physical substrates.
(2) Physical substrates and what they do (processes) are multiple realizable.
(3) When a substrate is inactive:
    (a) no self is present; and
    (b) substrate properties remain.
4) Therefore, selves are likely the substrate activities (what they do/the processes).

1) Selves are correlated to substrates. This appears to be the core of both the mind–body problem and the persistence problem. Within human history where a self is, there is a correlating body. There is persistent occurrence (here the thesis does not address the possibility of spiritual entities), i.e., where the person or the mind goes, so goes the body (this is the core of the animalism argument as discussed in chapter 3). Therefore, let it be assumed that selves (minds and persons however defined) require some substrate to be instantiated in. For the mind–body problem (even for the interactive substance dualist), there is some connection between minds and bodies and the core of the problem is in explaining how this may be. For the physicalist (whether standard physicalist property substrate or further physicalist property substrate, or a combination thereof), this means there is a physical body and that, somehow, the mind is a result of how this physical body hangs together (processes). Therefore, a physicalist of any kind may prefer the term 'instantiated' to 'correlated'. Whichever term is preferred, there is a general consensus among physicalists that the physical substrate interacts in some way (whether internally, externally, or some combination thereof) that results in a mind.

The persistence problem also affirms this correlation/instantiation, with the biological solution going so far as to say they are identical. For the current thesis, 1) is a reaffirmation of the physicalist stance by asserting that in order for there to be a physical process there needs to be a

physical substrate to instantiate it. However, the thesis distances itself from a view that only a particular substrate (the numeric body substrate) can instantiate a process and is optimistic that, as science develops, the processes of the mind may be replicated in alternative substrates.

2) Physical substrates and what they do are multiple realizable. If a substrate is physical (however defined) and, therefore, instantiates physical processes (those that occur within the substrate and in relation to the environment), then it is likely that an artifact could, in principle, replicate the substrate and the processes of said substrate. This is the argument for multiple realizable physicalism (see section 2.2.), which has been demonstrated to be congruent with multiple physicalist stances throughout chapter 2.

3) When a substrate is inactive, a) no self is present, and b) substrate properties remain. Consider an inactive substrate, such as a corpse. This is not the partial cessation of some processes (e.g., sleep), but rather the case is that all the known processes at a particular nomological level have stopped at some point.

First consider a), where no self is present when there are no processes (see the corpse problem presented in section 3.5.). The primary mind–body problem for the physicalist is how a mind can be the result of physical processes and, given that bodies can exist without a mind when a corpse, it can be inferred that processes of some kind are necessary for a mind. For the persistence problem, there is an inherent assumption that persons are dynamic, as without the changes there would be no problem that arises. Therefore, both problems accept that there are some processes occurring and that these processes are somehow related to selves (minds and persons). Following the common assumption that when processes cease, selves cease (a corpse is not a self), it can be stated that processes are essential to selves. Furthermore, the empirical trajectory of multiple fields (e.g., neuroscience, psychology) continues to explore the correlates of biological activity (what the substrate does) with psychological experiences (this includes aspects such as behaviour, embodied cognition, neural activity, and so on). In essence, where there are no processes there are no selves.

Now consider b), where we have an inactive substrate (e.g., a corpse) but no selves (as discussed above) and this substrate still has many of the substrate properties (e.g., weight, length, anatomical structure, space-time continuity, and so on). It, therefore, seems likely that these properties (the inactive properties of the property-process distinction this thesis discussed earlier) are not essential to the self. If no self is present yet certain properties may remain, these properties cannot be sufficient for selves. The current thesis makes the assertion that the primary distinction between the corpse and the self is, therefore, that it is the processes that matter, and this leads to the conclusion 4).

4) Selves are likely the substrate activities. If the self can only occur when there are processes and substrates can occur without selves, it is likely that the self is the processes as this is the differentiating condition in these scenarios. Alternatively, it could be inferred that some further aspect (e.g. a soul) is present with the substrate as well as the processes when the self is present. However, the principle of parsimony would indicate that processes would then be the preferred identity claim for a self rather than positing a further entity.

The process-self view, therefore, offers a way to explain and explore what the metaphysical nature of the self is as well as unifying multiple physicalist stances. Furthermore, the process-self view explains why the mind–body problem and the persistence problem emerged, in that it states there is a metaphysical difference between the mind/person and body, where the body is the inactive properties of a substrate and the mind/person being what that substrate does. The idea of the nomological boundary may further act as a means to determine which processes (Hl, Ll, environment, or any number of levels in between or beyond these levels) are likely to be appropriate for constituting a process-self and welcomes empirical research to further clarify these issues. This is not to say that the process-self view is without problems of its own, but the claim here is that it offers a way of exploring the problems that could potentially lead to greater solutions.

In relation to the MUP, the process-self view's identity claim that the self (minds and persons) *is* the processes of the substrate, would allow for all forms of upload scenarios (gradual or scan-copy, destructive or non-destructive). It has been shown in the previous chapter (section 3.5.) that the same process can be stopped, restarted in alternative substrates, as well as continue in a particular substrate. Thus, the identity claim of process-selves does not lie with the substrate that instantiates the processes, but with the processes themselves. The next section now turns to one possibility of what these processes could be and offers the current thesis' preferred stance to the concept of a self.

4.3 The efferent-self

Whereas the previous sections within the current chapter have aimed to unify multiple stances and, therefore, allow for multiple avenues for the MUP to potentially pursue, the current section focuses on a novel approach to the concept of a self (here meaning the processes of the mind/person), namely the concept of the efferent-self. Certain emphases are made and areas of contention that still exist to the author are mentioned. The aim here is not to present an exhaustive, well-developed theory on the self, but to present a brief sketch of what the current thesis suggests to be the preferred approach.

To begin, the current section presents some clarifications on where within the MUP map (see section 4.1.) the efferent-self stands. First, the self is deemed to be instantiated in a standard physical substrate and is multiple realizable. An area that is unclear, is the precise relation between standard physical processes (Ll) and the Hl processes (here a representational process). The current thesis holds to a weak emergent view (as opposed to a strong emergent or panpsychist view), but also carries the intuition that the nomological boundary may be drawn at the Hl levels. Similar to the carpenter who may have a theory of woodwork without knowledge of atomic structure (despite the weak emergence view that woodwork is nevertheless causally reducible), the current thesis holds that the Hl processes are likely to be nomologically autonomous in relation to artifact creation. As the relation between Ll and HL is unclear at present, the current thesis awaits further empirical research to establish at what precise level the nomological boundary, or boundaries, may be drawn. A further intuition is that there will be multiple meso-levels that are yet to be established. It is unclear as to whether Hl-processing is itself causal or whether it is rather an *ad hoc* categorising of Ll neural activities. The view of the current thesis is that Hl processes are causal, but do not form a new property. The Hl processes are, therefore, not deemed to be of a different metaphysical nature than the Ll processes, but to be a variation of complexity and function (similar to the flight of single bee being different to the flight of the same bee in a swarm, yet it is still flight).

Second, the standard physical substrate of the efferent-self emphasises the neurocentric view (as opposed to embodied cognition and pure behaviourist views) but, nevertheless, includes bodily processes and external processes (how this may be understood is developed further below).

Third, the concept of the efferent-self is a variation of the concept of the process-self (see previous section) and affirms the psychological solution to the persistence problem. In relation to personal identity, the concept of the efferent-self holds to qualitative identity (processes are deemed as qualities that can be continued in alternative substrates) that is primarily intrinsic, allows for replication (branching) with space-time instantiations, and affirms persons as type-persons (see chapter 3 and section 4.1.2.). The result in relation to the MUP is that these preferences would allow for any of the upload scenarios (e.g., gradual destructive upload, non-destructive scan-copy, and so on).

Having asserted that the self is a process (process-self), the current thesis presents what more accurately may be asserted as the necessary and sufficient processes for the self. The thesis now presents its own novel variation of the concept of the process-self, termed the concept of the efferent-self. The basic claims are:

a) There are efferent and afferent processes needed for a self to exist but it is the efferent processes that are directly related to the self.
b) The self is, furthermore, essentially a particular type of efferent process (representations and their interactions) that, in turn, refers (the nature of representations is to refer) to all other forms of efferent and afferent processes.

a) There are efferent and afferent processes needed for a self but it is the efferent processes that are directly related to the self. First, what is meant by efferent and afferent processes? The term efferent in its simplest form can be expressed as that which is conducted outwards (i.e., outputs) from a given system. In its turn, the term afferent relates to that which moves towards a system, (i.e., inputs) and, therefore, there is an implication of spatial location embedded in the term. Although this is commonly the case in living organisms (and adopted by the MUP scenarios emphasised in this thesis of a singular physical substrate occupying space-time), the meaning of the term may also be used to refer to a process of control. That the self is a control system of sorts is one of the primary qualities of persons as identified in chapter 3 and is accepted by the view of the efferent-self developed here. In a representational control system (to be considered further below), the efferent (what outputs extend from the system) may be localised or distributed. Consider a software program that is run from different computers in different locations across the world. In this scenario the efferent does not extend from one space-time location but from multiple space-time locations, yet the processes are efferent as they extend from the system[50]. In the efferent-self, the concepts of efferent and afferent are broadly defined and may be localised (e.g., in the brain) or distributed (e.g., as described in the software program above).

The substrate uses these efferent and afferent processes to interact with the environment (however defined). In the current human substrate, efferent and afferent processes are conducted through the nervous system; within a robot there may be other substrate parts (e.g., different types of afferent sensors) that perform these processes. Consider the nomological domain of cognitive neuroscience. Within neuroscience, afferent and efferent axons (parts of the biological substrate) convey information (the process) from the brain (efferent) and those that relay information to (afferent) the brain (Kandel, *et al.,* 2013). For example, when a person walks, the nervous system

50 The brain may be modularly localised (e.g., Carruthers 2004), where specific brain parts have specific functions or distributed (e.g., Anderson's radical redeployment hypothesis, 2007). What is interesting in relation to the MUP, is that, even if modularised locally, precise locations may vary (e.g., consider neural plasticity). In both of these options, it is the functions that matter, whether localised or distributed within the nervous system.

needs to coordinate/control this process. The brain sends out information (efferent) instructing the body what to do, but it is also receiving information (afferent) such as the pressure of the ground on the foot, orientation to other objects, and so on. To coordinate movement, both efferent and afferent processes are needed. When walking, both afferent and efferent processes are largely non-conscious, although attention may be drawn to some of these processes. For example, many of the efferent processes involved in walking may fall outside of awareness (e.g., which neurons fired to produce the effect), but some may also be brought into awareness (e.g., awareness of the movement of limbs in the walking process). The assertion of the efferent-self view is, therefore, not that a self needs to be aware of all efferent and afferent processes, but rather that both are necessary for any coordination of the self. Furthermore, the brain is not sending only one efferent or processing input from one afferent process at any given time. Therefore, for any phenomenon there may be multiple efferent and afferent processes as well as variation in ability to integrate both processes[51].

Consider the experiment where a person's arm is connected to a machine that is connected to another person's brain (Gage, n.d. – presented in section 3.5). In this experiment, the person whose arm is being manipulated sends out no efferent processes (is told not to think about moving their arm), whereas the person whose brain is connected to the machine initiates efferent processes (thinks about moving the arm). The person whose arm moves, receives afferent signals of this movement (e.g., proprioception stimuli) and reports that it is not *me/I* that is moving the arm. Another person/self is declared to be moving the arm (the person whose brain is connected to the machine). In this instance, the afferent processes of the arm's movement supersedes the efferent processes (no signal is sent to move the arm) and this discrepancy appears to be sufficient for the designation of a self not being attributed or as not being responsible (control as a necessary and sufficient conditions for the self). In functionalist terms, the self-system has inputs and outputs (occurring at multiple levels) and the efferent processes are the sum of the output processes and the afferent relates to the sum of the inputs. The body, when it enacts efferent processes, is attributed to the self in terms of actions of the self (the self-referencing aspect is to be discussed further under the second premise of representations), whereas when the processes are primarily afferent, these are denied as being actions of the self.

---

51 Substrates may vary in their ability to receive stimuli (e.g., may vary in their afferent processes). This variation may occur within a substrate type (such as persons with differing abilities in seeing, hearing, and so on), within the same substrate (e.g., an ability, such as hearing, may be lost), as well as across substrate types (e.g., some birds can see ultraviolet light, whereas humans cannot). This difference in afferent abilities also holds true for efferent abilities (e.g., birds can fly).

A further example that illustrates distinctions of the self from the non-self through efferent and afferent processes can be text document software, such as that which the current thesis is being typed on. If the software responds as expected (the software is functioning normally), I can make the statement that I am typing (the efferent processes supersede the afferent processes), but what would happen if the software program were to malfunction and the software generated random letters and symbols when I pressed the keys? Would that be my writing? The efferent-self view would attribute the processes that are under my control (when the software is functioning normally) as mine, whereas the other scenarios (where the afferent processes dominate) would be attributed to the software (the efferent processes are those of the software). The concept of the efferent-self provides a way of attributing the self in many scenarios and asserts that when there are efferent processes, the self may be attributed to those efferent processes, as the efferent-self.

The idea that afferent and efferent processes are needed, yet only the efferent processes are attributed to the self, may relate back to the concept of necessary but not sufficient conditions. For example, oxygen is a necessary condition for the self in the current biological substrate. If oxygen is removed, there can be no self (as all afferent and efferent processes would cease). But oxygen cannot be equated with the self, as it is not sufficient for the self (oxygen exists without the self). Similarly, afferent processes occur in many causal systems[52] (many substrates receive afferent processes), but it is only when there are efferent processes from the substrate (including a particular kind of efferent process namely, representational processes to be discussed below) that a self could possibly be said to be present. This is the distinction between animate (efferent and afferent) and inanimate objects (only afferent), of which life may be described as one form of animate object.

Efferent (and afferent) processes occur at multiple levels, such as external processes and internal processes. For example, the external processes of behaviour are efferent processes, in that they are processes under the control of the self, extend from the self, and are the direct result of efferent neural processes. Yet, behaviours are adaptive to the environment and, therefore, need to include afferent processes to develop and be maintained. Internal efferent processes of the self may include the sum of all efferent processes (at the level of the entire nervous system) but can also include efferent and afferent processes at lower levels, such as a specific neuron firing. When a

52 Cause and effect could be described as afferent (cause) and efferent (effect), but this definition would be too broad and include all forms of physical interactions. Furthermore, the question of where the initial cause should be attributed (should we include the big bang?) is also complex. Efferent processes are, therefore, defined here as relating to those that proceed from the control system (however defined) and, furthermore, asserts that all open dynamic systems (such as the person is considered to be, see chapter 3) include causes that fall outside of what may be termed the self.

single neuron fires, there is the stimulation of the neuron (afferent to the neuron) and the firing of the neuron sending an efferent process down the neuron to the synaptic cleft. This neuron would then act as afferent process to the next neuron. Therefore, there are afferent and efferent processes occurring at multiple levels, and decisions of nomological boundaries would need to be made as to what level of the efferent processes should be related to the self (to be discussed further under the second claim).

However, that efferent processes are to be associated with the self does not necessarily mean that selves are metaphysically all efferent processes. Is a single neuron firing, or a single cell organism's behaviour (both which have afferent and efferent processes) sufficient to justify the identity claim for a self? What if an efferent signal is started by another person as in the example above of the person who controls the arm of another through a machine? Or more simply, consider a doctor testing the reflexes of a patient where they hammer the nerve below the knee-cap and an efferent process is initiated so that the leg of the patient kicks? An efferent process is initiated outside of the brain of the self in all these scenarios and, therefore, not all efferent processes may be included in the concept of a self.

So, the self needs a way to distinguish between efferent processes of the self, efferent processes initiated by the other, and afferent processes, as well as to coordinate/control and adapt the efferent processes as the self interacts with its environment. However, it appears that coordination of efferent processes is also not sufficient. For example, a simple insect (e.g., a cricket), or a modern robot, may have efferent and afferent processes as well as processes of coordination, but these are not deemed by the efferent-self view as sufficient for a mind (if one does, consider the moral implications of humanities' current actions to these 'selves'). The efferent processes are, therefore, not any cause (afferent) and effect (efferent) system that can be coordinated but must include something beyond simple coordination. How are the efferent-afferent processes distinguished in selves and how is this coordination and adaptation achieved? What types of coordination, or what else, is needed for a self to be present? A precise answer is yet to be elucidated, but the efferent-self view makes the assertion that the notion of a representational system may offer appropriate solutions and this is the second premise of the efferent-self-view. The concept of the efferent-self includes efferent processes (a), but further asserts that the efferent processes are controlled by a representational system (b).

b) The self is, furthermore, a particular type of efferent process (representations and their interactions) that relates to all other forms of efferent and afferent processes in turn. The efferent-self view is not an identity theory stating that all efferent processes (such as an efferent axon being stimulated) are identical with the self, but rather a representational theory in asserting that what is

the self is the representation of these efferent processes. Consider the development of a child. At first, the child enters the world and is bombarded with stimuli (e.g., afferent sights, sounds, textures) that it needs to make sense of and coordinate behaviour around these stimuli[53]. To do this, the child develops representations (within developmental psychology these are often termed schema – see e.g., Piaget, 1952). But the child is also developing a sense of self and establishing a distinction between the self and the world (Sugarman & Jaffe, 1990; Rochat, 2003), such as in identifying the self in the mirror. What the efferent-self view posits, is that the child is developing a representation of the self (including the body and what the self controls – see Gallagher, 1995), and to do this, it relies on efferent and afferent processes (i.e., both efferent and afferent processes are needed to develop such representations).

Precisely how the distinction between that which is efferent and that which is afferent may be achieved, is still contentious but the efferent-self view, at this junction, aligns itself with predictive coding (e.g. Gladziejewski, 2016, Metzinger & Wiese, 2017). The general idea (and its variants such as predictive processing) as it relates to mental phenomena, is that there is a prediction (the function of the brain is to form a hypothesis both consciously and non-consciously, and as these hypotheses are products of the brain, they are efferent processes) and there is feedback (afferent processes), such as from the environment. In this view, the brain's function is to minimise prediction error and acts as a form of Bayesian inference. The efferent-self view posits that efferent processes are those that are likely to have less prediction error and the afferent greater prediction error and, in taking this into account, the brain can compute what is attributed as efferent and what is afferent and, therefore, what should be attributed to the self and what should be attributed to the non-self. It is beyond the scope of the current thesis to explore what is a burgeoning field of enquiry, which includes various perspectives that could operationalise further research (e.g., Friston's free energy principle and application of Markov blankets as a means to distinguish entities, and so on – Friston, 2010, Kirchoff *et al*, 2018; also see Metzinger and Wiese, 2017).

For the concept of the efferent-self then, the primary metaphysical nature of selves (minds and persons) is taken to be a representational control system that creates representations of the body

53 The efferent-self view allows that certain representations are 'hard-wired' into the genetics (e.g., see Evolutionary Psychology views such as Barkow, Cosmides, & Tooby, 1992) and accepts that both nature and nurture are needed for the development of a mind. The claim of the efferent-self view is, therefore, not an argument as to when these representations are formed, but simply just that a human mind is a representational system (however these representations originated).

(including the brain) and the environment, or even imagined environments, and that it is the representations of efferent processes that constitute the self. To clarify, the claim here is that the representations of efferent processes are the self and not that all the efferent processes are the self (although as stated above within the efferent-self view efferent processes are associated with the self by the representational system). Here a complication arises as to the nature of representations. The efferent-self account holds to the view that all representations are efferent processes themselves (currently being instantiated in the brain) that refer to something. Representations, in this view, are efferent because they are the product of the physical processes of the brain and are here termed 'repeffs' (representational efferent processes). It is these repeffs that distinguish the efferent processes of a cricket from the efferent processes of a self. The repeffs may refer to afferent processes (a representation of inputs), non-representational efferent processes (representation of outputs), or to themselves (repeffs of repeffs).

To explore the notion that mental phenomena are not identical to bodily processes (whether afferent or efferent) but rather to the representations of the bodily processes, consider the condition of an injury that severs the nerves in the spine such as a paraplegic injury (one could also consider local anaesthesia or any other situation, where the afferent processes do not reach the brain). The mental phenomenon of pain cannot be experienced in such cases, because even if there is stimulation of the nervous system in the limb (afferent process), no representation (the efferent process) is activated. Pain is a complicated philosophical issue and used as an example in multiple philosophical stances as presented in chapter 2. Here it is noted that the concept of pain presents the analytic philosopher (here meaning philosophy that emphasises common language) with some difficulties, because pain is a phenomenon that is, at times, deemed a part of the self and, at others, as not part of the self. For example, when a pain is experienced the common phrasing is such that "I have pain. I am in pain" and not "I am paining. I pain". These common phrasings distinguish the self from the pain (pain is what happens to me and is, therefore, not me). Yet, pain is still my mental phenomenon and, therefore, forms a part of me somehow.

Now consider the case of phantom limbs, where pain is being experienced in a limb that does not exist (e.g., Flor *et al*., 1995). If pain were directly related to only afferent processes, then how can the person experiencing phantom limb pain gain such an experience? In essence, how can such a pain exist without the afferent processes (there is no limb for the pain to be sent from)? The answer may lie in the assertion that all mental phenomena are repeffs in the brain. On the efferent-self account, pain is a representation that is produced by (efferent) the brain and afferent processes (here the stimulation of the nervous system such as C-fibres and the like) relate to what is the non-self. Because the representations are efferent, these are what are considered to be part of the self

(what is my repeff of afferent processes is my pain). For the repeff, it does not matter whether there is an actual afferent process being instantiated, but only that there is repeff of that afferent process. Yet the pain is still not me (I may have a pain but the pain is not me), because it is usually an afferent process and afferent processes are commonly deemed by the self to not be part of the self. Therefore, the efferent-self account offers a way to understand how pain can be the non-self (when it is the repeff of afferent processes) and still be part of the mental phenomena of the self (as it is a repeff).

This line of reasoning is similar to Smart (1959 – see section 2.5.2.1) who put forward that the phenomenal first person experiences (e.g., WIL, qualia, phenomenal properties), such as the experience of seeing colour, is a representation (an aggregate of past neural processes). A representation, if so understood, is a predictable pattern of past neural patterns produced by (efferent) the brain. In this sense, the efferent-self view is an identity theory in some sense (representations are identical to neural patterns) but not in a 'bodily' sense (pain is not C-fibre stimulation but the representation of C-fibre stimulation).

Furthermore, repeffs, as variations of the notion of the process-self (see above), are physical processes and not entities that instantiate processes. The assertion is that there can be no representation without physical processes and that representations aren't things but processes. That these representations may appear as 'things' can be explained from the perspective of the efferent-self account as a category mistake, which Place (1956) termed the phenomenological fallacy (see section 2.5.2.1), where there is a confusion over phenomenal properties as things when, in fact, they are processes (e.g., neural activity). A repeff may be real (scientifically validated) as the physical processes are real, but what a repeff refers to may not be real. For example, in the repeff of a unicorn, the referent of unicorn is not real but the processes that produce the representations are (repeffs are standard physical processes).

Therefore, the efferent self-account asserts that all representational processes are real (i.e., in humans the neural processes and in standard computing the activities of the hardware), however, not all referents are real. The distinction between the repeff process as a real physical processes (the ontological assertion of physicalism) and the referent as ontologically questionable, may explain some of the difficulties in current philosophy of mind debates. For example, the mind may have a referent of the self that is not real (e.g., a delusion that I am Napoleon or that I have magic powers), that is real (e.g., the neural processes), refers to real objects in the world (e.g., trees, dogs, and so on), as well as imaginary objects (e.g., unicorns). And so, the efferent-self account maintains a distinction between the repeffs (the physical processes) as real, whereas what they refer to (whether to themselves or to the outside world) may not be.

When repeffs refer to other repeffs (a neural process refers to another neural process), there is a "strange loop" (Hofstadter, 1979), in that the representational system is representing itself. As such, the self is a meta-cognition, a higher-level representation (although the account of the efferent-self does not deem these to be of a distinct metaphysical order as the higher order theories do – presented in section 2.6.2). The self is, therefore, a self-referencing indexical system, where the self is a representation of itself. This self-representation also refers to recursive dynamic processes, as efferent processes change over time as the self interacts and adapts to the environment. In the efferent view, the self is, therefore, self-indexing and dynamic.

Returning to the previously stated options of repeffs either referring to afferent processes (a representation of inputs), non-representational efferent processes (representation of outputs), or to themselves (repeffs of repeffs). Repeffs of the world and others refer to afferent processes, repeffs of the overall body refer to partly afferent and partly efferent processes, and repeffs of repeffs refer to purely efferent processes that, according to the efferent self-view, occur only within the brain[54]. The account of the efferent-self views these as a spectrum from highly predictable and controllable with minimal prediction error (the repeffs) to less predictable and controllable (afferent). Efferent processes may, therefore, be placed on a spectrum from more efferent (where the self has primary control) to less efferent (in which case, afferent processes are also involved). The self is deemed to have an attribute of control and, therefore, repeffs are to a greater extent the self than non-repeff efferent processes and, furthermore, afferent processes (of which there is even less predictability). Non-repeff efferent processes (such as the signals sent to a limb) require that the brain processes afferent processes alongside it (e.g., when walking). This inclusion of afferent processes into the repeff coordination system increases the possible prediction error and, therefore, decreases the likely allocation to the self. In contrast, a pure repeff process (such as imagination) has very limited afferent processes at the nomological level of representations and their interaction and is, therefore, more attributable to the self (less need to process minimisation error).

As stated earlier, afferent and efferent processes may occur at multiple levels and this may be further asserted in relation to repeffs. Consider a standard computer that is a representational

54 The efferent-self account does not claim that all brain activity is efferent, nor does it claim that all brain activities are representations but, given the current limitations of our knowledge, simply claims that the brain is an appropriate organ to further explore these possibilities. There is also the possibility of embodied representations (Svensson & Ziemke, 2005), which would indicate that repeffs may be expanded to include the body, although this is not the view of the efferent-self account as presented within the current thesis.

system (although it is believed to have different representational processes from a brain). The desktop, with its representations of icons, can be further reduced to various levels of symbolic (representational) processes until the level of binary representations (0 or 1 attributions). The account of the efferent-self acknowledges that selves interact with multiple levels of representations to produce the overall representation of the self. The efferent-self is a unifying concept that is a meta-representation of the efferent processes (both repeffs and non-repeff efferent processes). The concept of the efferent-self, therefore, includes multiple levels of efferent processes, with repeffs being the most efferent (i.e., they have less interference from afferent processes). In terms of functionalism, the efferent-self is the representation of the sum of the outputs of a given system. Precisely how repeffs are distinguished from non-repeff efferent processes is yet to be established. Repeffs may involve a greater number of neurons (be more largely distributed) or may be primarily be the function of a specific neural network. Repeffs may take more time to be instantiated and so may be seen as more reflective than reflexive. The 'how' of the repeff falls beyond the scope of the current section, which aims to do no more than to briefly sketch the distinctions between repeffs (representational efferent processes) and non-repeff efferent processes.

What the account of the efferent-self also presents, is an approach that allows for the self to both be extended and retracted and, in so doing, may provide possible avenues to explore borderline cases (such as the experiment where a person's arm is moved without their control, phantom limb pain, psychiatric pathology such as schizophrenia, and so on) and provide a distinction between a forensic self and a metaphysical self (on personal identity as a forensic self-see Locke, 1689; Nimbalkar, 2011). For the advocate of the efferent-self view, the forensic self (i.e., that persons are morally responsible for their actions) is the efferent processes (the behaviours and so on that extend from the representational system). All efferent processes that are included by the repeff system as the actions of the system (here assuming some form of prediction error as the mechanism by which this is achieved) are deemed to be the self. For example, the movement of the body in striking someone with your fist is a forensic offence, as it is if a similar efferent processes is enacted with tool use such as a blunt object. In these instances, the repeffs send out a control signal and the body, or the tool extension, enacts these efferent processes and, so, is deemed to be part of the forensic self.

However, if an efferent process occurs without the correlating repeffs (e.g., when another system initiates the efferent process, such as a doctor testing a patients reflexes), this is not deemed by the repeff system as an act of the self and, therefore, it is not the self that is performing the

actions[55]. The concept of a metaphysical self is, therefore, not simply the efferent processes but includes that the repeffs (including the self-repeff as the unifying representation of all efferent processes) are deemed to be in control of these efferent processes.

Precisely where the lines are to be drawn between the forensic self and the metaphysical self remains to be established. The efferent spectrum has been suggested earlier as a way to establish the more efferent from the less efferent but requires greater clarity. For example, at which part of the system does the repeff become a non-repeff (a non-representational efferent signal)? Is it the brain as a whole, or part of the brain, or could it relate to signal degeneration over the nervous system? The current thesis awaits further empirical and philosophical research in this regard. The efferent-self account, therefore, allows for all efferent processes to be attributed to the self (forensic self) but denies that these efferent processes are themselves the metaphysical nature of the self (only repeffs are the essential self). The concept of the forensic self allows for attributions of extension to include the body (beyond the nervous system) and tool use and may use the notion of prediction error as a means to make these attributions of extension to the self. The efferent-self can be said to extend through efferent processes.

From the view of the efferent-self, the self may be deemed a fluid dynamic process where the self extends or retracts. In relation to retraction, consider the person who loses a limb (or even temporarily loses the use of a limb, such as when someone falls asleep in an awkward position and the limb is numb on awakening). Loss of a limb function does not amount to a loss of a self as the efferent-self (the sum of all efferent processes coordinating processes) still continues. If the overall repeff system continues to function, then the self continues. Even if most common body efferent processes are limited, such as in persons who have locked-in syndrome, the self continues so long as the repeffs continue. Also, retraction need not be limited to the body and may include the brain. Consider a memory (a representation) that once served to be part of the causal processes but now is forgotten and serves no efferent role. For the efferent-self, what serves no efferent process is no

55 A complication may arise when psychiatric pathology is present (e.g., delusions of control in schizophrenia) and a crime is committed. A subcategory of psychology is forensic psychology (e.g., Tredoux, et al., 2005), where it may be queried whether at the time of the offence the accused was responsible (capable of controlling their own actions). For the efferent-self view, this may be explored through the assertion that, although the act was committed with efferent processes (e.g., efferent behaviour), these are the result of a fractured self (the self-repeff does not acknowledge these efferent processes as part of the self). Therefore, further work needs to be done to establish a sufficient level of coordination of repeffs for a self to be responsible for their actions.

longer part of the self and, therefore, such memories that were once part of the efferent-self are no more and the efferent-self has been retracted.

It could, however, be argued that the forgotten memory still has some causal role to play in the repeff system even though it is non-conscious one. This would mean that all memories are continuously causal even if they can never be retrieved. As a thought experiment, consider that a memory could, in principle, be removed from one's brain. Not only can't the memory be retrieved, but the part of the brain that once stored that memory (or however one can conceive of performing this thought experiment) is removed. Would it be said that the self no longer exists? If the intuition is that the self continues without this memory, then it can be said that some repeffs that may have been part of the repeff system may be removed and, therefore, self-retraction is possible. The intuition of the current thesis is that much of the self may be retracted without the self-ceasing to exist, and that the psychology of personal continuity is more robust than some would assume.

Returning to the self as extendable. The efferent-self is similar to Clark and Chalmer's notions of an extended mind (1998 see section 2.5.1.3), in that both consider the self to be a computational/ representational process, although the concept of the efferent-self has a different emphasis, namely the efferent processes. Consider extension as mentioned above, such as tool use. When cutting an onion, it is common to say that I cut the onion, yet technically it is the knife that performed the incision. The efferent-self confirms the use of 'I' in this context, as there are greater efferent processes transferred through the medium of the knife and there are fewer afferent processes (the feel of the knife cutting the onion is felt through the medium of the knife). Most inanimate objects that one exerts control over (where there are more efferent processes) are, therefore, deemed to be extensions of the self (see previous comments on the forensic self).

For the extended mind theory, tool use (ibid.) is a more radical claim, in that the mind is not merely an extension of the efferent processes (such as in the use of tools) but the tool becomes part of the mind (the mind itself extends). The efferent-self view is, therefore, a more frugal hypothesis that allows for extension of the self through tools but restricts the metaphysical self to repeffs. If the tool itself does not produce repeffs (Otto's book may produce a representation but the status of the book as an efferent processes may be queried), the tool may be integrated within the repeff system as an extension of the self (the forensic self) but not of the essential metaphysical self (which is limited to repeffs). However, this does not mean that the repeff system could not be extended. Consider a future computer interface was made to be a continuous part of causal processes (e.g., a person with locked in syndrome may use an artificial intelligence computer that generates its own repeffs). In this future scenario the computer is part of the repeff system and, therefore, the efferent-self view would deem the mind to be extended.

What has been presented here is a brief sketch of a novel idea, the notion of an efferent-self. It aligns with many views, such as those asserting that selves are instantiated in standard physical substrates, it is a functionalist (although it emphasises efferent processes as the necessary and sufficient aspects of the self), representational, and computational, view of the mind, and so on. As such, it is susceptible to the same criticisms that these views have received (e.g., the Chinese room thought experiment, the knowledge argument, and so on). In relation to the MUP, the concept of the efferent self presents a view of the self that would readily allow for the MUP in all scenarios. First, it is the psychological processes (here viewed as repeffs) that matter for the self to continue. Second, these processes are not limited to a particular substrate, and third, the emphasis on representation allows for the possibility of both biological and artificial artifact substrates (given the appropriate computational properties).

4.4 Summary of the integrative chapter

What this chapter has aimed to achieve, is to integrate many of the philosophical stances and views in relation to the MUP. The first section focused on integrating the two primary premises for the metaphysical feasibility of the MUP, namely; that selves need to be the instantiated by multiple realizable physical substrates and that a psychological solution to the persistence problem is upheld. The second section emphasised that the process-self view may be a way to explore both the mind–body problem and the persistence problem. Furthermore, this section demonstrated that a focus on processes allows for the MUP in many different upload scenarios. Finally, the chapter presented the section on the account of the efferent-self as a novel approach to the concept of the self. Advancing from the account of the process-self, the account of the efferent self asserts that it is efferent processes and, more specifically, the repeffs that are to be the preferred emphasis in relation to selves.

5 Conclusion

A self-awakes with their mind, their thoughts, their memories, their desires, their beliefs. Is this person you? This is what happened to the reader, the author, and every other person this morning. If this offers no problem for our identification of ourselves, let us consider what happens when we turn the dials of our thinking on this issue. For example, imagine the same scenario but with part of your body missing, is it still you? What if your entire body was missing but the mind (wherever you think the mind is located, or whatever you think its nature is) remains, is it still you? What if part of your mind is missing, such as a memory, is this still you? What if part of the brain is missing, is it still you? What if that part of the brain is replaced with an artifact, is this still you? What happens if the entire brain is replaced, is it still you? The core concept of the MUP is that 'you' will continue irrespective of what substrate is active, so long as your mind (however defined) is instantiated in that substrate. The current thesis, therefore, enquired into the metaphysical nature of 'you' (persistence of personal identity over time) and the metaphysical nature of the mind (developed in the context of the mind–body problem) that would allow for 'you' to continue across substrates.

The metaphysical nature of the mind was the subject of chapter 2 and established that the MUP is metaphysically feasible should multiple realizable physicalism be maintained as a necessary and sufficient condition. The metaphysical nature of persons was the subject of chapter 3, where the psychological solution to the persistence problem as a means of personal identity was established as the second necessary and sufficient metaphysical condition for the MUP.

Throughout the current thesis, it has been affirmed that for the MUP, which aims to develop an artifact that will instantiate a particular mind in an alternative substrate, the emphasis is on the substrate and what the substrate does. This has been related by the current thesis to Sellars' (1963) discussion of the aim of philosophy as exploring 'things' (entities/substrates) and 'how things hang together' (processes/what the substrate does). For the MUP to be successful, minds (chapter 2) and persons (chapter 3) cannot be the 'thing'/substrate (i.e., the continuity of the particular substrate and its particular parts in space-time) but rather minds and persons must be the continuity of what the substrate does. If these are the same processes (as has been the assertion of the current thesis), then a continuity of these processes are both the continuity of the mind and the person. In essence, persons are minds and whatever substrate instantiates that particular mind instantiates that particular person.

If the same mind is said to be successfully instantiated in an alternative substrate, it would be classified as a 'mind upload' (the mind occurs across substrates). In this view, if minds are replicable, then so are persons, as the identity claim defended in this thesis in chapter 3 is that a

person is their mind. If the nature of a mind were such that the mind cannot be replicated (instantiated in an alternative substrate), then the MUP would be non-feasible.

This led to the subject of chapter 2, which queried what is the nature of the mind and is it such that it can be instantiated across substrates? In this chapter, the necessary and sufficient conditions for a mind to be successfully instantiated across substrates was discussed within the context of the mind–body problem. Considering the distinction between substrates and what substrates do, the current thesis initially explored the mind–body problem by considering whether the substrate is physical (physicalism) or non-physical (emphasising substance-dualism, see section 2.4.1). A non-physical substrate would interact and produce non-physical processes (what the substrate does), whereas a physical substrate (however defined) would produce physical processes. Any philosophy (e.g., substance dualism) that asserts that the physical and non-physical interact, would need to explain how this interaction does or could occur (this is the primary problem of mental causation and over-determination). The current thesis argued that the substrate is physical (however defined) and that, therefore, what the substrate does are physical processes. These physical processes were then further categorised as occurring within the substrate (internal processes) or occurring outside of the substrate (external processes). The physical substrates were then further categorised as being constituted by standard physical properties (section 2.5.) or as further physical properties (section 2.7.), with the further physical properties being mental properties beyond (further) standard physical properties (e.g., mass, energy, and so on), but nevertheless taken to be somehow physical.

The MUP requires that the mind be replicable though the human efforts of artifact design and creation. Any artifact produced is, by definition, the result of necessary and sufficient knowledge for said replication and the necessary and sufficient techniques and tools to produce the artifact. The type of knowledge and methods for artifact creation fall within the overarching domain of science. The current thesis has, following Sellars (1963), emphasised (initially presented in section 1.5.7. and developed throughout chapter 2) the scientific image (as opposed to the manifest image) as the preferred epistemology from which metaphysics can be developed and therefore argued, that there is a natural alliance between scientific research and the metaphysical research of this thesis (viewed as possible trajectories and probable metaphysics based on current science). In essence, in relation to the MUP, for the mind to be replicable, it would need to be of the nature that would allow it to potentially fall within the purview of some current or future science.

The overarching domain of science may be divided into nomological domains that reflect different areas of scientific enquiry (e.g., quantum physics, atomic theory, biology, psychology, sociology, and so on). One of the novel ideas of the current thesis (initially presented in section

1.5.6. and developed throughout chapter 2), is to acknowledge that for artifact creation a nomological boundary or boundaries may be drawn (a designation of necessary and sufficient conditions for replication in the context of the thesis) to determine which domains are necessary and sufficient. This nomological boundary may be applied whether the domain under consideration is reducible (e.g., the solidity of wood that the carpenter works with) or non-reducible (e.g., as property dualists assert in relation to the mind). The notion of a nomological boundary has been applied within the current thesis as it provides, irrespective of the stance on reducibility, clarity on the necessary and sufficient conditions for replication.

The emphasis of the nomological boundary within this thesis has primarily been drawn in terms of external (e.g., behaviourism) internal processes (e.g., identity theory), with the internal processes being further categorised into lower levels (e.g., neuroscience) and higher levels (e.g., psychology). The use of the term 'levels' aligns, although does not limit, the current thesis to some form of emergence. The thesis presented Hl (higher levels) as synonymous with psychology (the mind), which relates to a phenomenon that emerges from standard physical properties (section 2.5, 2.6, and 2.7.4) or it may be a parallel property, such as in panpsychic property dualism (section 2.7.1). The Hl options as presented by chapter 2 and summarised in section 4.1.1 (see Table 4-1), are that Hl (the mind) may either be identified directly with standard physical processes (Ll = Hl), may be non-causal but the result of standard physical processes (Ll →Hl), may be reducible to standard physical properties but be the product of emergence such as in weak emergence (Ll ←→Hl), may not be reducible to but be the result of standard physical processes as in strong emergence (Ll ←→Hl), or may not be reducible to standard physical properties because the Hl consist of some further physical properties, such as in panpsychic property dualism (Ll ←→Hl). Furthermore, the Hl may be an autonomous nomological domain (the current thesis emphasised information processing as one such possibility in section 2.6.2), where autonomy may be reducible (e.g., weak emergence) or non-reducible (strong emergence, or panpsychic property dualism). For artifact creation, it is not essential that the reducibility or non-reducibility be resolved (such as the carpenter example), but that the necessary and sufficient nomological boundary or boundaries for artifact creation have been identified and sufficiently understood. This allows for multiple philosophical stances, each with their own assertions where nomological domains are necessary and sufficient, to be allowed in the context of the MUP.

What has been primarily argued in chapter 2 is that, irrespective of how the Hl is understood (e.g., Hl as an alternative description of Ll, Hl in terms of non-reductive panpsychic property dualism, and so on), that so long as the properties and processes at Hl are physical (of whatever ilk) and multiple realizable, then the MUP is feasible and that this, therefore, acts as the primary

necessary and sufficient condition for the MUP. The current thesis has asserted that the properties and processes of the mind (if they are attributed to physical, of standard or further physical, properties and processes) are a) within the realm of science and, therefore, of artifact creation and b) multiple realizable.

Multiple realizability has been broadly defined by the thesis in terms of allowing for the possibility of a property or process being instantiated in an alternative substrate[56]:

(1) The mind is a series of physical processes or properties (however defined).
(2) Physical processes and properties are likely to be multiply realizable.
(3) Therefore, the mind is likely to be multiply realizable.

The idea of the multiple realizability in relation to the mind has been explored throughout the current thesis (initially presented in section 2.2 and developed throughout chapter 2) and here a brief summary is presented by considering the above argument. The first premise is the assertion of physicalism (here meaning physical monism that allows commitment to either standard physical properties and processes or further physical properties and processes). As such, it denies forms of non-physical substrates and substrate processes (e.g., idealism and substance dualism, see section 2.4) which, if true, would indicate that the mind is beyond the purview of any current or future science (science being the enquiry into the physical domain) and, therefore, beyond the purview of artifact creation sufficient for the MUP.

The second premise is an inference based on the empirical sciences, named by the current thesis as the empirical trajectory (initially presented in section 1.5 and developed throughout chapter 2). First consider the notion of properties. Scientifically established physical properties (e.g., mass, energy, atomic particle properties, and so on) belong to multiple substrates and, when appropriately understood, potentially may be used to produce an artifact that instantiate the same properties (here assuming that a standard physical substrate is the necessary and sufficient condition). Should a further physical substrate be needed, the empirical trajectory (as discussed in section 2.2 and 2.7.) indicates that these further physical mental properties are also likely replicable. In addition, the type sortal spectrum can be applied to properties in asking whether the property type may

56 The reader is reminded that the notion of multiple realizability in multiple realizability physicalism (see section 2.2.1) is not limited to realizations occurring across substrate kinds (e.g., pain occurring in octopuses and humans) or within the same substrate (e.g., different pain occurring within the same person), but relates to any repeatable phenomenon.

have multiple tokens or whether it is an absolutely unique type. As all known physical property types may have multiple physical tokens (e.g., mass, energy, and so on occurs in multiple tokens), these physical property types may be understood as types with multiple tokens and, therefore, relate to multiple realizable physicalism (recognising the possibility of properties that occur across substrates). As all known physical properties are multiple realizable, in this way it can be inferred that all mental properties (whether via weak or strong emergence, panpsychism, or any further physicalist variant) will also be multiply realizable. The alternative that there may be a property that is unique in the universe to each particular person is deemed less likely (see sections 1.5.5 and 3.6).

Furthermore, given then that properties are multiple realizable, it can be inferred that physical processes are multiple realizable. When properties are sufficiently understood (as per the applicable nomological boundary), the processes that occur when these properties interact are likely replicable (e.g., chemical properties resulting in a particular reaction when processes are initiated, resulting in change in colour). Both properties and processes as understood within multiple realizable physicalism relate to the nature of the scientific method and technological replication, where experiments and artifact creation are established through repeatable observations of the same effect and this reproduction/replication/multiple instantiations of a phenomenon are what leads to scientific knowledge (broadly understood) (see Section 2.2). In this sense, difficulties of process replication may relate to engineering difficulty rather than a metaphysical one. For example, weather systems are difficult to predict and would, therefore, be difficult to replicate. However replication of a weather system is a matter of engineering complexity rather than the process being, in essence, non-multiple realizable (weather being a physical system). This complexity may affirm that there is a pragmatic limitation to artifact creation but does not necessarily imply that there is a metaphysical limit.

If all the appropriate properties are present and the appropriate processes instantiated (both internally and externally), then the same phenomenon can, in principle, be instantiated. Throughout chapter 2 it was demonstrated that many of the physicalist stances place an emphasis on processes such as the external processes of the behaviourist, the neural processes of the identity theorist, the functional processes of the functionalist, and so on. From this, the current thesis then further asserted that the phenomenon of the mind is a process (what the substrate does) rather than property (what the substrate is) (see section 4.2). From the assertion that the notions of mind and persons are metaphysically a process (as developed throughout chapter 2 and further developed in the process-self view in section 4.2), it was noted that processes may be moved across substrates (such as the half and half slinky example in section 2.6., where a slinky continues a wave function from a plastic slinky substrate to a metal slinky substrate) as well as stopped and started in alternative

substrates (e.g., the domino cascade example presented in section 3.3), provided similar enough properties are present (the substrate properties) and appropriately integrated (e.g., neural circuitry similarity for the mind).

The thesis then turned to the persistence problem in chapter 3, which considered what the necessary and sufficient conditions for persons are to continue across substrates. As persons change over time, the question is how the same personal identity may be retained and what conditions are needed for this continuity? The thesis argued that absolute continuity is not feasible in relation to persistence of persons (the problem of persistence emerges through the acknowledgement that a person is different over time) and that a notion of partial continuity should be embraced.

The only account of persons that appears to relate to non-replication is the one that emphasises the substrate continuity in space-time as the necessary and sufficient condition for persons (the biological solution). This may be related to a notion of numeric identity (section 3.5) and a notion of non-branching (section 3.7). In the personal continuity thought experiments (section 3.4), we were asked to imagine a molecule for molecule transfer (we could extend this as quantum string for quantum string, if so inclined), where a replica has no distinguishable features/attributes other than the original, except for the substrate not having continuity in space-time. The biological solution to persistence of identity would have us believe that this one feature is the cornerstone of personal identity. Thus, the claim is that when this one attribute is removed, all other qualitative claims of identity (that the person is precisely the same in every other respect) are insufficient for persistence of identity.

The current thesis has attempted to show that this claim is false and that it is the notion of qualitative identity that is more likely the condition for continuity of persons (with an emphasis on the qualities of the mind). The qualities that are deemed necessary and sufficient conditions by the current thesis are the qualities of the mind, thus those espoused by supporters of the psychological solution to the persistence problem. Furthermore, by exploring certain notions, such as space-time instantiations, dynamic-open systems, indexical representational systems, and control systems, the thesis presented a way of understanding the intuition that supports the biological solution, while denying that the numeric attribute (space-time continuity of substrate) this solution advocates for is, in fact, what metaphysically matters (section 3.5).

The current thesis also explored other options (see chapters 3 and chapter 4), such as a numeric psychological solution, branching/non-branching in alternative substrates, persons as types with potentially alternative tokens, and so on. In all of these instances, once the assertion that

space-time continuity of the particular substrate is denied as essential for continuity of persons/minds and the psychological solution is maintained, the MUP becomes a metaphysically feasible project.

Having presented the mind as multiple realizable (chapter 2) and persons as minds (the psychological solution of chapter 3), the thesis integrated these in chapter 4 using the term 'self' as the unifying concept under which minds (the subject of chapter 2) and persons (the subject of chapter 3) could be integrated. Chapter 4 first (section 4.1) explored how the assertions of multiple realizable physicalism and the psychological solution to persistence of identity lead to the feasibility of the MUP, and how these assertions may relate to the various stances within philosophy of mind. From there, chapter 4 presented (section 4.2) the concept of the process-self, where the identity claim of the self should be best understood as the continuity of a process. In this view, the self is what the substrate does (processes). The chapter (section 4.3) then considered what kind of processes are best identified with the self and presented a more refined process-self view, that of the efferent-self. Within the potential nomological boundaries that relate the substrate (the body) to the processes of the substrate (the mind), the concept of the efferent-self identifies the mind as a representational system that identifies the self with efferent processes. Efferent processes are processes that proceed from the system (outputs) and are contrasted with afferent processes as those processes that are received by the system (inputs). Representations and their interactions are also efferent processes (termed repeffs by the current thesis) and, so, are included in the self. The claim of the efferent-self view is that most efferent processes are commonly identified by the self with the self but that it is only the representational efferent processes (repeffs) that are actually (metaphysically) the self. The notion of the efferent-self is currently a fledgling idea and requires further philosophical and empirical work.

The current thesis has left many questions unanswered. From a metaphysical perspective, the feasibility of the MUP may be challenged from many non-physicalist perspectives (e.g., soul theories, substance dualism, idealism) and, within physicalism, there is still the possibility (although deemed unlikely by the current thesis) that minds may require a non-replicable (i.e., non-multiple realizable) property to be instantiated. These options have been acknowledged as contra-indicating the MUP (the C of the ABC options initially introduced in section 1.3 and developed primarily in chapter 2) but due to the limitations of the current thesis, may still require further exploration.

Furthermore, there are many engineering questions that remain unanswered. For example, at which nomological level will the nomological boundary be the necessary and sufficient condition for the MUP? Although the thesis has presented its own current preference (Hl-processing), this

remains to be established and awaits further research. Even if the Hl-processing view is affirmed as the appropriate nomological boundary, there are still many areas for further clarity and exploration as to how the self is actually developed and maintained. For example, further research is needed to determine whether predictive coding offers appropriate processes, whether a binary Turing computer can perform the process, whether quantum computing may be needed, and so on.

The current thesis has, therefore, enquired into the necessary and sufficient metaphysical conditions for mind uploading and argued that multiple realizable physicalism and the psychological solution to persistence of identity are two such conditions. If these conditions are upheld, then the MUP is feasible. Furthermore, the assertion that the self is best understood as a process, presents a way to unify disparate philosophical stances in relation to the MUP and resolve some potential problems. For example, much of the mind–body problem rests on the disparate features of the mental and the physical (see section 2.3) and the notion of the process-self asserts that this disparity is based in the distinction between substrates (things) and what substrates do (processes). In relation to the persistence problem, the process-self also resolves the problem of the continuity of persons, in asserting that it is continuity of processes and not continuity of substrates that matter. The notion of the efferent-self presents a further way forward that would, if successful, provide further solutions and avenues for the MUP (e.g., relating efferent processes to predictive coding) that would allow for any upload scenario.

6 Bibliography

Akand, M. (2018). Identity theory, multiple realizability of mental states and functionalism. *Pabna University of Science and Technology Studies*, *3*, 79–86. Retrieved from https://www.researchgate.net/publication/338535390_Identity_theory_Multiple_Realizability_and_Functionalism

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders *(5th ed.). Washington, DC: American Psychiatric Publishing.*

Anderson, M. L. (2007). The massive redeployment hypothesis and the functional topography of the brain. *Philosophical Psychology*, *20*(2), 143–174. doi: 10.1080/09515080701197163

Anscombe, G. E. M. (1975). *The first person*. Oxford, UK: Clarendon Press.

Andrade, G. (2018). Philosophical difficulties of mind uploading as a medical technology. *APA Newsletters*, *18*(1), 208–214. http://0-search.ebscohost.com.innopac.wits.ac.za/login.aspx?direct=true&db=pif&AN=EP133172232&site=eds-live&scope=site

Armstrong, D. M. (1968). *A materialist theory of the mind*. London, UK: Routledge & Kegan Paul.

Armstrong, D. M. (1999). The open door: Counterfactual versus singularist theories of causation. In Sankey, H (Ed.) *Causation and laws of nature* (pp. 175–185). New York, NY: Springer. doi: 10.1007/978-94-015-9229-1_16

Astakhov, V. (2008). Mind uploading and resurrection of human consciousness. Place for science? *NeuroQuantology*, *6*(3). 245-261. ISSN 13035150. doi: 10.14704/nq.2008.6.3.181

Audi, R. (Ed.). (1999). *The Cambridge dictionary of philosophy* (Second edition). Cambridge, UK: Cambridge University Press.

Ayer, A. J. (1936/1962). *Language, truth, and logic*. London, UK: Victor Gollanze.

Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.

Baker, L. R. (2000). *Persons and bodies: A constitutional view*. Cambridge, UK: Cambridge University Press.

Baker, L. R. (2005). When does a person begin? *Social Philosophy & Policy*, *22*(2), 25–48.

Barkow, J. H., Cosmides, L., & Tooby, J. (1992). *The adapted mind*. New York, NY: Oxford University Press.

Barrett, D. A. (2013). Multiple realizability, identity theory, and the gradual reorganization principle. *The British Journal for the Philosophy of Science*, *64*(2), 325–346. doi: 10.1093/bjps/axs011

Bateson, G. (1979). *Mind and nature: A necessary unity*. New York, NY: Dutton.

Benedikter, R., Siepmann, K., & Reymann, A. (2017). “Head-transplanting” and" mind-uploading": philosophical implications and potential social consequences of two medico-scientific utopias. *Review of Contemporary Philosophy*, *16*, 38–82. doi: 10.22381/RCP1620172

Berkeley, G. & Bennett, J. (1713/2017). *Three dialogues between Hylas and Philonous in opposition to sceptics and atheists*. Retrieved from https://www.earlymoderntexts.com/assets/pdfs/berkeley1713.pdf

Berkeley, G., & Williams D.R. (1734/2002). *A treatise concerning the principles of human understanding*. London, UK: Jacob Tonson. Retrieved from https://www.maths.tcd.ie/~dwilkins/Berkeley/HumanKnowledge/1734/HumKno.pdf

Bickle, J. (1992). Multiple realizability and psychosocial reductionism. *Behavior and Philosophy, 20*(1), 47–58.

Bickle, J. (1996). New wave psychophysical reductionism and the methodological caveats. *Philosophy and Phenomenological Research*, *56*(1), 57–78. doi: 10.2307/2108465

Bickle, J. (2019). Multiple realizability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019). Stanford, CA: The Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2019/entries/multiple-realizability/

Bitchel, Wi. & Mundale, J. (1999). Multiple realizability revisited. *Philosophy of Science*, *66*(2), 175–207. doi**:** 10.1086/392683

Blatti, S. (2012). A new argument for animalism. *Analysis*, *72*(4), 685–609. doi: 10.1093/analysis/ans102

Blatti, S. & Snowdon, P. F. (2016). Introduction. In Blatti, Stephan & Snowdon, Paul. F. (Eds.), *Animalism: New essays on persons, animals, and identity*. (pp. 1-27). Oxford, UK: Oxford University Press.

Blatti, S. & Snowdon, P. F. (2016) (Eds.), *Animalism: New essays on persons, animals, and identity*. Oxford, UK: Oxford University Press.

Block, N. (1980). Troubles with functionalism. In Block, Ned (Ed.), *Readings in philosophy of psychology* (Vol. 1, pp. 268–305). Harvard University Press, Cambridge, MA.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. doi: 10.1017/S0140525X0003847

Block, N. (1996a). Mental paint and mental latex. *Philosophical Issues*, *7*, 19–49. doi: 10.2307/1522889

Block, N. (1996b). *What is functionalism?* The Encyclopedia of Philosophy Supplement. Pp 27-44 Retrieved from https://www.researchgate.net/publication/28762327_What_Is_Functionalism

Block, N. (2008). Phenomenal and access consciousness Ned Block and Cynthia MacDonald: Consciousness and cognitive access. *Proceedings of the Aristotelian Society*, *108*, 289–317. Retrieved from JSTOR database. doi: 10.1111/j.1467-9264.2008.00247.x

Bobro, M. (2017). Leibniz on Causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017). Retrieved from https://plato.stanford.edu/archives/fall2017/entries/leibniz-causation/

Born, M. (1949). Natural philosophy of cause and chance. Oxford UK, Clarendon Press.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford, UK: Oxford University Press.

Bostrom, N., & Sandberg, A. (2008). *Whole brain emulation: A roadmap* (Technical Report No. 2008–3; p. 2015). https://www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf

Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe? *Philosophical Studies*, *170*(3), 465–500. doi: 10.1007/s11098-013-0259-7

Braidotti, R. (2013). *The posthuman*. Cambridge, UK: Polity Press.

Brentano, F. (1995). *Descriptive psychology* (Müller, B. Trans.). London, UK: Routledge.

Brown, J. R. (1995). Thought experiments. *Canadian Journal of Philosophy*, *25*(1), 135–142. Retrieved from https://www.jstor.org/stable/pdf/40231903.pdf?refreqid=excelsior%3A99eff06cd0834ef1d91b09a5e4273c0d

Brown, J., Huntley, D., & Morgan, S. (2018). Confabulation: Etiology, typology, and intervention. *Clinical Research in Neurology*, *1*(1), 1–4. Retrieved from https://asclepiusopen.com/clinical-research-in-neurology/volume-1-issue-1/7.php

Burge, T. (1979/2002). Individualism and the mental. In D. J. Chalmers (Ed.), *Philosophy of mind: Classic and contemporary readings* (pp. 597–607). New York, NY: Oxford University Press.

Burge, T. (2007). *Foundations of mind* (Vol. 2). Oxford, UK: Oxford University Press.

Bužek, V., & Hillery, M. (1996). Quantum copying: Beyond the no-cloning theorem. *Physical Review A*, *54*(3), 1844–1852. doi: 10.1103/PhysRevA.54.1844

Campbell, J. (2018). Philosophy Overdose. (n.d.). *5 Functionalism & Putnam – Berkeley Lectures.* Retrieved from https://www.youtube.com/watch?v=AVlMa4VdwhQ&t=2s

Campbell, S (2006). The conception of a person as a series of mental events. *Philosophy and Phenomenological Research*, *73*(2), 339–358. doi: 10.1111/j.1933-1592.2006.tb00621.x

Candland, D. K. (1993). *Feral children and clever animals: Reflections on human nature*. New York, NY: Oxford University Press.

Cappuccio, M. L. (2017). Mind-upload. The ultimate challenge to the embodied mind theory. *Phenomenology and the Cognitive Sciences*, *16*(3), 425–448. doi: 10.1007/s11097-016-9464-0

Carnap, R. (1991). Logical foundations of the unity of science. In Boyd, Richard, Gasper, Philip, & Trout, J. D. (Eds.), *The Philosophy of Science* (pp. 393–404). Cambridge, MA: The MIT Press.

Carruthers, P. (2004). The mind is a system of modules shaped by natural selection. In Hitchcock, Christopher (Ed.), *Contemporary debates in the philosophy of science* (pp. 293–311). Malden, MA: Blackwell Publishing, Ltd.

Carruthers, P. (2016). Higher-Order Theories of Consciousness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2016). Retrieved from https://plato.stanford.edu/archives/fall2016/entries/consciousness-higher/

Cerullo, M. A. (2015). Uploading and the branching identity. *Minds & Machines*, *25*, 17–36. doi: 10.1007/s11023-014-9352-8

Chakravartty, A. (2017). Scientific Realism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Retrieved from https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219. Retrieved from http://personal.lse.ac.uk/ROBERT49/teaching/ph103/pdf/chalmers1995.pdf

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford, UK: Oxford university press.

Chalmers, D. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies*, *17*(9-10), 7-65. Retrieved from http://consc.net/papers/singularity.pdf

Chalmers, D. J. (2014). Uploading: A philosophical analysis. In *Intelligence Unbound: The Future of Uploaded and Machine Minds.* R. Blackford & D. Broderick (eds). Wiley, Blackwell: Sussex, UK. pp. 102-118.

Chalmers, D. (2015). Panpsychism and panprotopsychism. In T. Alter & Y. Nagaswa (Eds.), *Consciousness in the physical world: Perspectives on Russellian monism* (pp. 246–276). Oxford, UK: Oxford University Press.

Churchland, P. M. (1981). Eliminative materialism and propositional attitudes. *The Journal of Philosophy*, *78*(2), 67–90. doi: 10.5840/jphil198178268

Churchland, P. S. (1986). Neurophilosophy: Toward a unified science of the mind/brain. Cambridge, MA: MIT Press.

Churchland, P. M., & Churchland, P. S. (1997). Recent work on consciousness: philosophical, theoretical, and empirical. *Cognitive Studies*, *4*(3), 45–55. doi: 10.11225/jcss.4.3_45

Churchland, P. S., & Sejnowski, T. J. (2016). *The computational brain* (25th Anniversary). Cambridge, MA: A Bradford Book.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. doi: 10.1093/analys/58.1.7

Copeland, B. J., & Shagrir, O. (2018). The Church-Turing thesis: Logical limit or breachable barrier? *Communications of ATM*, *62*(1), 66–74. doi: 10.1145/3198448

Copi, I. M., & Cohen, C. (1980). *Introduction to logic* (Eighth edition). New York, NY: Macmillan Publishing Company.

Corabi, J., & Schneider, S. (2012). Metaphysics of Uploading. Journal of Consciousness Studies *19* (7): 26-44. Retrieved from http://sjupsych.org/corabi/documents/Metaphysuploading.pdf

Daffertshofer, A., Plastino, A. R., & Plastino, A. (2002). Classical no-cloning theorem. *Physical Review Letters*, *88*, 210601-1-210601-4. doi: 10.1103/PhysRevLett.88.210601

David, D., Cristea, I., & Hofmann, S. G. (2018). Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Frontiers in Psychiatry*, 9: 4. doi 10.3389/fpsyt.2018.00004

Davidson, D. (2001). Mental events. In D. Davidson's *Essays on actions and events*. (pp. 207-225) Oxford, UK: Clarendon Press. doi:10.1093/0199246270.003.0011

Davidson, D. (2013). Knowing one's own mind. The American Philosophical Association Centennial Series, 389–409. doi: 10.1093/0198237537.003.0002

Dawkins, R. (2003). *The devil's chaplain*. New York, NY: Houghton Mifflin Company.

Dawkins, R. (2011). *The magic of reality*. London, UK: Bantam Press.

Dennett, D. C. (1978a). Current Issues in the Philosophy of Mind. *American Philosophical Quarterly*, *15*(4), 249–261. Retrieved from https://www.jstor.org/stable/i20009721

Dennett, D. C. (1978b). *Brainstorms: Philosophical essays on mind and psychology*. Cambridge, MA: Bradford Book, The MIT Press.

Dennett, D. C. (1981). Three kinds of intentional psychology. In R. Healey (Ed.) *Reduction, Time and Reality* (pp. 37–61. Cambridge, UK: Cambridge University Press. Retrieved from dl.tufts.edu/pdfviewer/vt150w469/fj236d81w

Dennett, D. (1982). Where am I? In D. R. Hofstadter, Douglas, R., & D. Dennett (Eds.), *The mind's I* (pp. 217–241). New York, NY: Bantam Books.

Dennett, D. C. (1988a). Quining qualia. In Marcel, A., & Bisiach, E. (Eds.), *Consciousness in contemporary science* (pp. 42–77). Oxford, UK: Oxford University Press. Retrieved from https://ase.tufts.edu/cogstud/dennett/papers/quinqual.htm

Dennett, D. C. (1988b). Précis of the intentional stance. *Behavioral and Brain Sciences*, *11*(3), 495–505. doi: 10.1017/S0140525X00058611

Dennett, D. C. (1989). *The intentional stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (1991). *Consciousness explained*. London, UK: Penguin Books.

Dennett, D. C. (1992). The self as a center of narrative gravity. In Kessel, F., Cole, P., & Johnson, D. (Eds.), *Self and consciousness: Multiple perspectives*. (pp. 103–115) Retrieved from http://cogprints.org/266/1/selfctr.htm. doi: 10.5209/rev-ASEM.2013.v46.42862

Dennett, D. C. (2013). *Intuition pumps: And other tools for thinking*. London, UK: Penguin Books.

Dennett, D. C. (2017). From bacteria to Bach and back: The evolution of minds. London, UK: Penguin Books.

Descartes, R. (1637/1968). *A discourse on method: Meditations and principles*. (Veitch, J., Trans.). New York, NY: J.M Dent & Sons.

Dresler, M., Wehrle, R., Spoormaker, V. I., Koch, S. P., Holsboer, F., Steiger, A., & Czisch, M. (2012). Neural correlates of dream lucidity obtained from contrasting lucid versus non-lucid REM sleep: A combined EEG/fMRI case study. *Sleep*, *35*(7), 1017–1020. doi: 10.5665/sleep.1974

Dretske, F. (1988). *Explaining behaviour*. Cambridge, MA: MIT Press.

Dreyfus, H. L. (1972). *What computers can't do*. New York, NY: Harper & Row Publishers.

Efron, R. (1970). The minimum duration of a perception. *Neuropsychologia*, *8*(1), 57–63. doi: 10.1016/0028-3932(70)90025-4

Eliasmith, C. (2013). How to build a brain: A neural architecture for biological cognition. Oxford, UK: Oxford University Press. doi: 10.1016/0028-3932(70)90025-4

Emmeche, C., Køppe, S., & Stjernfelt, F. (1997). Explaining emergence: towards an ontology of levels. *Journal for General Philosophy of Science*, *28*(1), 83–117. doi: 10.1023/A:1008216127933

Endicott, R. P. (1993). Species-specific properties and more narrow reductive strategies. *Erkenntnis*, *38,* 303–321. doi: 10.1007/BF01128233

Evans, G. (1978). Can there be vague objects? *Analysis*, *38*(2), 208. doi: 10.1093/analys/38.4.208

Feigl, H. (1958/2002). The "mental" and the "physical." In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings*. New York, NY: Oxford University Press.

Ferrari, G. R., & Griffith, T. (2000). *Plato: The Republic*. Cambridge, UK: Cambridge University Press Cambridge.

Feyerabend, P. K. (1963). Mental events and the brain. *The Journal of Philosophy*, *60*(11), 295–296. doi: 10.2307/2023030

Flor, H, Elbert, T., Knecht, S., Wienbruch, C., Pantev, C., Birbaumer, N., Larbig, W., & Taub, E. (1995). Phantom-limb pain as a perceptual correlate of cortical reorganization following an arm amputation. *Nature*, *375*(6531), 482–484. doi: 10.1038/375482a0

Fodor, J. A., (1974). Special sciences and the disunity of science as a working hypothesis', Synthese, 28: 77–115. doi: 10.1007/BF00485230

Fodor, J. A. (1975), *The Language Of Thought*. Sussex, UK: The Harvester Press, Ltd.

Fodor, J. A. (1987). *Psychosomantics*. Cambridge, MA: MIT Press.

Forrest, P. (2020). The Identity of Indiscernibles. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/identity-indiscernible/

Frankish, K. (2007). The anti-zombie argument. *The Philosophical Quarterly*, *57*(229), 650–666. doi: 10.1111/j.1467-9213.2007.510.x

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, *23*(11–12), 11–39. Retrieved from https://nbviewer.jupyter.org/github/k0711/kf_articles/blob/master/Frankish_Illusionism%20as%20a%20theory%20of%20consciousness_eprint.pdf

Freud, S. (1954). Project for a scientific psychology. In M. Bonaparte, A. Freud, & E. Kris (Eds.), & E. Mosbacher, J. Strachey, E. Mosbacher, & J. Strachey (Trans.), *The origins of psychoanalysis: Letters to Wilhelm Fliess, drafts and notes: 1887–1902.* (pp. 347–445; By S. Freud). doi: 10.1037/11538-013

Freud, S. (1955). *Psychological works of Sigmund Freud* (Stanford edition). London, UK: Hogarth Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. doi: 10.1038/nrn2787

Gage, G. (n.d.). *How to control someone else's arm with your brain*. Retrieved July 18, 2020, from https://www.ted.com/talks/greg_gage_how_to_control_someone_else_s_arm_with_your_brain

Gallagher, S. (1995). Body schema and intentionality. In Bermudez, J., Eilan, N., & Marcel, A. J. (Eds.), *The body and the self* (pp. 225–244). Oxford, UK: Oxford University Press.

Gallois, A. (2016). Identity over time. In E. N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy* (Winter 2016). Stanford, CA: The Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2016/entries/identity-time/

Gazzaniga, M. S. (1995). Principles of human brain organization derived from split-brain studies. *Neuron*, *14*(2), 217–228. doi: 10.1016/0896-6273(95)90280-5

Gazzaniga, M. S. (2005). Forty-five years of split-brain research and still going strong. *Nature Reviews Neuroscience*, *6*(8), 653–659. doi: 10.1038/nrn1723

Geach, P. (1973). Ontological relativity and relative identity. In M.K. Munitz (ed.). Logic and Ontology, New York, NY: New York University Press.

Gell-Mann, M. (1994). The quark and the jaguar: Adventures in the simple and the complex. New York, NY: Holt Paperbacks.

Gibbard, A. (1975). Contingent identity. *Journal of Philosophical Logic*, *4*(2), 187–221. doi: 10.1007/BF00693273

Gillett, C. (2002). The metaphysics of realization, multiple realizability, and the special sciences. *The Journal of Philosophy*, *100*(11), 591–603. doi: 10.2307/3655746

Gladziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, *193*(2), 559–582. doi: 10.1007/s11229-015-0762-9

Goertzel, B. (2012). When should two minds be considered versions of one another? *International Journal of Machine Consciousness*, *4*(01), 177–185. doi: 10.1142/S1793843012400094

Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Science, 13*(4), 154-159. doi: 10.1016/j.tics.2009.01.007

Gould, S. J. (1997). The exaptive excellence of spandrels as a term and prototype. *Proceedings of the National Academy of Sciences*, *94*(20), 10750–10755. doi: 10.1073/pnas.94.20.10750

Grandy, R. E. (2016). Sortals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2016/entries/sortals/

Gregory, R.L. (Ed.). (1998). *The Oxford companion to the mind*. Oxford, UK: Oxford University Press.

Griffin, N. (1974). *Relative identity* (Doctoral dissertation). The Australian National University, Australia. http://hdl.handle.net/1885/10438

Hahn, L. E. (1999). *The Philosophy of Donald Davidson*. Chicago, IL: Open Court Publishing.

Hampshire, D. P. (2018). A derivation of Maxwell's equations using the Heaviside notation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2134):20170447, 1–13. doi: 10.1098/rsta.2017.0447

Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives*, *4*, 31–52. doi: 10.2307/2214186

Harnish, R.M., & Cummins, D.D. (eds.) (2000) *Minds, brains, and computers*. Hoboken, NJ: Wiley.

Hauskeller, M. (2012). My brain, my mind, and I: some philosophical assumptions of mind-uploading. *International Journal of Machine Consciousness*, *4*(01), 187–200. doi: 10.1142/S1793843012400100

Hayworth, K. (2010). Killed by bad philosophy. The Brain Preservation Foundation. https://www.brainpreservation.org/content-2/killed-bad-philosophy/

Heil, J. (2012). Philosophy of mind: A contemporary introduction. New York, NY: Routledge.

Hempel, C. G. (1949/1980). The logical analysis of psychology. *Readings in Philosophy of Psychology*, *1*, 14–23.

Hempel, C. G. (1958). The theoretician's dilemma: A study in the logic of theory construction. *Concepts, Theories, and the Mind-Body Problem*, *02*, 37–98. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/184621

Hobbes, T. (1651/1973). *Leviathan*. London, UK: J.M Dent & Sons.

Horgan, T. (1984). Functionalism and token physicalism. *Synthese*, *59*(3), 321–338. doi: 10.1007/BF00869338

Horgan, T. (1993). From supervenience to superdupervenience: Meeting the demands of a material world. *Mind*, *102*(408), 555–586. doi: 10.1093/mind/102.408.555

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid* (20th Century). London, UK: Penguin Books.

Hofstadter, D. R. (1982). Reflections (What is it like to be a bat?). In Hofstader, D. R., & Dennett, C. (Eds.), *The mind's I* (pp. 403–414). New York, NY: Bantam Books.

Hofstadter, D. R. (1982). A conversation with Einstein's brain. In Hofstadter, D. R., & Dennett, C. (Eds.), *The mind's I* (pp. 430-456). New York, NY: Bantam Books.

Huemer, W. (2019). Franz Brentano. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019). Stanford, CA: The Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2019/entries/brentano/

Hume, D. (1777/1975). Enquiries concerning human understanding and concerning the principles of morals (3rd ed.). Oxford, UK: Oxford University Press.

Huxley, T. H. (1874). On the hypothesis that animals are automata, and its history. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 24–30). Oxford, UK: Oxford University Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly (1950-)*, *32*(127), 127–136. doi: 10.2307/2960077

Jackson, F. (1986). What Mary didn't know. *The Journal of Philosophy*, *83*(5), 291–295. doi: 10.2307/2026143

Jackson, F., Pargetter, R., & Prior, E. W. (1982). Functionalism and type-type identity theories. *Philosophical Studies*, *42*(2), 209–225. doi: 10.1007/BF00374035

Jacob, P. (2019). Intentionality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Stanford CA: Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/intentionality/

Jaworski, W. (2011). *Philosophy of mind: A comprehensive introduction*. Sussex, UK: Wiley & Blackwell.

Jaworski, W. (2016). Structure and the metaphysics of mind: How hylomorphism solves the mind–body problem. Oxford, UK: Oxford University Press.

Jordan, M., Elsbernd, P., & Sladky, J. (2019). Thermoregulatory dysfunction in a patient with locked-in syndrome due to bilateral ventral pontine infarction (P4. 3-031). Minneapolis, MN: AAN Enterprises. Retrieved from https://n.neurology.org/content/92/15_Supplement/P4.3-031.abstract

Kaku, M. (2012). *Strings, conformal fields, and M-theory*. Berlin, DE: Springer Science & Business Media.

Kandel, E. R. (2007). In search of memory: The emergence of a new science of mind. New York, NY: WW Norton & Company.

Kandel, E.R., Schwartz, J.H., Jessel, T.M., Siegelbaum, S.A., & Hudspeth A.J. (Eds.). (2013). *Principles of neural science* (5th ed.). New York, NY: McGraw-Hill Companies.

Kant, I., & Meiklejohn, J.M.D. (Translator). (1781/1959). *Critique of pure reason*. London, UK: J.M Dent & Sons.

Keller, H. (1905). *The story of my life*. New York, NY: Grosset & Dunlap.

Kirchoff, M., Parr, T., Palacios, E, Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, *15*(138), 1–11. doi: 10.1098/rsif.2017.0792

Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research*, *52*(1), 1–26. doi: 10.2307/2107741

Kim, J. (2010). *Essays in the metaphysics of mind*. Oxford, UK: Oxford University Press.

Kim, J. (2011). *Philosophy of mind* (3rd ed). Boulder, CO: Westview Press.

Knight, W. (2018). MIT has just announced a $1 billion plan to create a new college for AI. Retrieved from https://www.technologyreview.com/the-download/612293/mit-has-just-announced-a-1-billion-plan-to-create-a-new-college-for-ai/

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*, *22*(3–4), 333–347. doi: 10.1080/02643290442000130

Kurzweil, R. (2005). *The Singularity is near: When humans transcend biology*. New York, NY: Viking.

Kripke, S. A. (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Larson, C. (2018). China's massive investment in artificial intelligence has an insidious downside. Retrieved from https://www.sciencemag.org/news/2018/02/china-s-massive-investment-artificial-intelligence-has-insidious-downside

Lau, J., & Deutsch, M. (2016). Externalism about mental content. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Stanford, CA: The Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2016/entries/content-externalism/

Lee, S. (2016). Occasionalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Stanford, CA: The Metaphysics Research Lab, Stanford University Retrieved from https://plato.stanford.edu/archives/win2016/entries/occasionalism/

Leek, J. T., & Peng, R. D. (2015). Opinion: Reproducible research can still be wrong: Adopting a prevention approach. *Proceedings of the National Academy of Sciences*, *112*(6), 1645–1646. doi: 10.1073/pnas.1421412111

Leibniz, G. W. V. (1965). *Monadology and other philosophical essays* (Schrecker, P., & Schrecker, A. M., Trans.). Indianapolis, IN: Bobbs-Merrill Educational Publishing.

Leiguarda, R., Starkstein, S., Nogues, M., Berthier, M., & Arbelaiz, R. (1993). Paroxysmal alien hand syndrome. *Journal of Neurology, Neurosurgery & Psychiatry*, *56*(7), 788–792. doi: 10.1136/jnnp.56.7.788

Levin, J. (2018). Functionalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Stanford, CA: The Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/fall2018/entries/functionalism/

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*, 354–361. doi: 10.1111/j.1468-0114.1983.tb00207.x

Lewis, D. K. (1966). An argument for the identity theory. *The Journal of Philosophy*, *63*(1), 17–25. doi: 10.2307/2024524

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, *50*(3), 249–258. doi: 10.1080/00048407212341301

Lewis, D. (1986). *On the plurality of worlds*. Oxford, UK: Oxford: Blackwell.

Lewis, D. K. (1987). Survival and identity. In *Philosophical Papers Volume 1* (pp. 55–72). Oxford, UK: Oxford University Press. Retrieved from http://home.sandiego.edu/~baber/metaphysics/readings/Lewis.SurvivalAndIdentity.pdf. doi: 10.1093/0195032047.003.0005

Lewis, D. (1988). What experience teaches. In *Papers in metaphysics and epistemology* (pp. 262–290). Cambridge, UK: Cambridge University Press. Retrieved from https://www.repository.cam.ac.uk/bitstream/handle/1810/247198/Lewis-Experience.pdf?sequence=2&isAllowed=y

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, *8*(4), 529–539. doi: 10.1017/S0140525X00044903

Libet, B., Freeman, A., & Sutherland, K. (Eds.). (1999). *The volitional brain*. Thorverton, UK: Imprint Academic.

Locke, J. (1689). *An essay concerning human understanding*. London UK: George Routledge and Sons Limited.

Loose, J., Menuge, A. J., & Moreland, J. P. (Eds.). (2018). *The Blackwell companion to substance dualism*. Hoboken, NJ: John Wiley and Sons.

Lowe, E. J. (2004). *An introduction to the philosophy of mind*. Cambridge, UK: Cambridge University Press.

Lowe, E. J., (2006). Non-Cartesian substance dualism and the problem of mental causation. Erkenntnis, 65 (1): 5–23. doi: 10.1007/s10670-006-9012-3

Ludlow, P., Nagasawa, Y., & Stoljar, D. (2004). There's something about Mary: Essays on phenomenal consciousness and Frank Jackson's knowledge argument. Cambridge, MA: MIT Press.

Lycan, W. G. (1998). The continuity of the levels of nature. In Lycan, W. G. (Ed.), *Mind and cognition: A reader* (pp. 77–96). Oxford, UK: Blackwell Publishers.

MacBride, F. (2016). Relations. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/relations/

Marshall, D., & Weatherson, B. (2018). Intrinsic vs. Extrinsic Properties. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2018/entries/intrinsic-extrinsic/

Matthews, W. J., Stewart, N., & Wearden, J. H. (2011). Stimulus intensity and the perception of duration. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 303-313. doi: 10.1037/a0019961

Mautner, T. (2005). *Penguin dictionary of philosophy* (2nd Edition). London, UK: Penguin Books.

McGinn, C. (1982). *The character of mind*. Oxford, UK: Oxford University Press.

McGinn, C. (1989). Can we solve the mind–body problem? *Mind*, *98*(391), 349–366. doi: 10.1093/mind/XCVIII.391.349

McGinn, C. (2007). The brain: An unbridgeable gulf. *Time Magazine*, *29*(January). http://content.time.com/time/magazine/article/0,9171,1580385,00.html

McLaughlin, B. (Ed.). (1991). *Dretske and his critics*. Cambridge, MA: Basil Blackwell.

McLaughlin, B., & Bennett, K. (2018). Supervenience. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2018/entries/supervenience/

McCullough, J. N., Zhang, N., Reich, D. L., Juvonen, T. S., Klein, J. J., Spielvogel, D., Griepp, R. B. (1999). Cerebral metabolic suppression during hypothermic circulatory arrest in humans. *The Annals of Thoracic Surgery*, *67*(6), 1895–1899. doi: 10.1016/S0003-4975(99)00441-5

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. doi: 10.1016/j.cortex.2010.11.002

Metzinger, T., & Wiese, W. (Eds.). (2017). *Philosophy and predictive processing*. MIND Group. Retrieved from https://predictive-mind.net/papers

Milekovic, T., Sarma, A. A., Bacher, D., Simeral, J. D., Saab, J., Pandarinath, C., Tringale, K. R. (2018). Stable long-term BCI-enabled communication in ALS and locked-in syndrome using LFP signals. *Journal of Neurophysiology*, *120*(7), 343–360. doi: 10.1152/jn.00493.2017

Minsky, M. (1988). *Society of mind*. New York, NY: Simon and Schuster.

Minsky, M. (2006). *The emotion machine*. New York, NY: Pantheon.

Moor, J. (2006). The Dartmouth College Artificial Intelligence conference: The next fifty years. *AI Magazine*, *27*(4). doi: 10.1609/aimag.v27i4.1911

Moore, J. T., Hren, C. R., & Mikulecky, P. J. (2015). *U Can: Chemistry I for dummies*. Hoboken, NJ: Wiley & Sons, Inc.

Moravec, H. (1988). *Mind children: The future of robot and human intelligence*. Cambridge, MA: Harvard University Press.

Morgan, C.L. (1927). *Emergent evolution* (2nd ed.). London, UK: Williams and Norgate.

Morgan, R. K. (2002). *Altered Carbon.* New York, NY: Penguin Random House.

Nagel, T. (1971/2008). Brain bisection and the unity of consciousness. In Perry, John (Ed.), *Personal identity*. pp. 227-248. Berkley, CA: University of California Press.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450. doi: 10.2307/2183914

Newman, L. (2019). Descartes' Epistemology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019). Stanford, CA: The Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2019/entries/descartes-epistemology/

Nida-Rümelin, M., & O Conaill, D. (2019). Qualia: The Knowledge Argument. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Stanford CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2019/entries/qualia-knowledge/

Nimbalkar, N. (2011). John Locke on Personal identity. *Mens Sana Monographs*, *9*(1), 268–275. doi: 10.4103/0973-1229.77443

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259. doi: 10.1037/0033-295X.84.3.231

Noonan, H., & Curtis, B. (2018). Identity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018). Stanford, CA: The Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2018/entries/identity/

Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.

Nichols, S., & Bruno, M. (2010). Intuitions about personal identity: An empirical study. *Philosophical Psychology*, *23*(3), 293–312. doi: 10.1080/09515089.2010.490939

Olson, E. T. (2004). Animalism and the Corpse Problem. *Australasian Journal of Philosophy*, *82*(2), 265–274. doi: 10.1080/713659837

Olson, E. T. (2007). *What are we? A study in personal ontology*. New York, NY: Oxford University Press.

Olson, E. T. (2017). Personal Identity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2017/entries/identity-personal/

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In *Minnesota studies in the Philosophy of Science* (Vol. 2, pp. 3–36). Minneapolis, MN: University of Minnesota Press.

Orilia, F., & Swoyer, C. (2020). Properties. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2020/entries/properties/

Papineau, D. (2 000). The rise of physicalism. In M. W. F. Stone & J. Wolff (Eds.), *Proper Ambition of Science* (pp. 174–208). London, UK: Routledge. doi: 10.4324/9780203446263

Papineau, D. (2002). *Thinking about consciousness*. Oxford, UK: Claredon Press.

Parfit, D. (1971). Personal identity. *The Philosophical Review*, *80*(1), 3–27. doi: 10.2307/2184309

Parfit, D. (1984). *Reasons and persons*. Oxford, UK: Oxford University Press.

Parfit, D. (2016). Divided minds and the nature of persons. In Schneider, Susan (Ed.), *Science Fiction and Philosophy* (2nd ed., pp. 91–98). Hoboken, NJ: John Wiley & Sons.

Park, J. L. (1970). The concept of transition in quantum mechanics. *Foundations of Physics*, *1*(1), 23–33. doi: 10.1007/BF00708652

Pavlov, I. P. (1927). Conditioned reflexes: An investigation into the physiological activity of the cerebral cortex (Anrep, G. V., Trans.). New York, NY: Dover Publications, Inc.

Peirce, C. S. S. (1906). Prolegomena to an apology for pragmaticism. *The Monist*, *16*(4), 492–546. doi: 10.5840/monist190616436

Penrose, R. (1989). *The emperor's new mind*. London, UK: Vintage.

Perry, J. (1976). The importance of being identical. In Amelie, Oksenberg, Rorty (Ed.), *The identities of persons* (pp. 67–90). Berkley, CA: University of California Press.

Perry, J. (Ed). (2008a). *Personal Identity* (2nd ed.). Oakland, CA: University of California Press.

Perry, J. (2008b). *Introduction* (Perry, John, Ed.; Second edition, pp. 3–30). Oakland, CA: University of California Press.

Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). New York, NY: International Universities Press.

Pigliucci, M. (2014). Mind uploading: A philosophical counter-analysis. In R. Blackford & D. Broderick (Eds.), *Intelligence unbound: The future of uploaded and machine minds* (pp. 119–130). Hoboken, NJ: Wiley & Blackwell.

Pitt, D. (2018). Mental Representation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018). Stanford, CA: The Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2018/entries/mental-representation/

Place, U. T. (1956). "Is consciousness a brain process?" *British Journal of Psychology*, *47*(1), 44–50. doi: 10.1111/j.2044-8295.1956.tb00560.x

Popper, K. R. (1953). Language and the body-mind problem. Proceedings of the 11th International Congress of Philosophy, 7, 101–107. doi: 10.5840/wcp1119537216

Popper, K., R. (1955). A note on the mind–body problem. *Analysis*, *15*. 131-135. doi: 10.1093/analys/15.6.131

Popper, K .R. (1983). *A pocket Popper*. (Miller, D., Ed.). Oxford, UK: Fontana Paperbacks.

Preston, J., & Bishop, M. J. (Eds.). (2002). *Views into the Chinese room: New essays on Searle and artificial intelligence*. Oxford, UK: Oxford University Press.

Prinz, J. J. (2012). *The conscious brain*. New York, NY: Oxford University Press.

Putnam, H. (1960). Minds and machines. In S. Hook (Ed.), *Dimensions of mind: A symposium* (pp. 148–179). New York, NY: New York University Press.

Putnam, H. (1965). Brains and behavior. In N. Block (Ed.), *Readings in Philosophy of Psychology* (Vol. 1, pp. 24–36). Cambridge, MA: Harvard University Press.

Putnam, H. (1975). The meaning of "meaning." *Language, Mind, and Knowledge*, *7*, 131–193.Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/185225.

Putnam, H. (1980). The nature of mental states. In N. Block (Ed.), *Readings in Philosophy of Psychology*, (Vol. 1, pp. 223–231). Cambridge, MA: Harvard University Press.

Quine, W. V. O. (1995). *From stimulus to science*. Cambridge, MA: Harvard University Press.

Quine, W. V. O. (2013). *Word and object*. Cambridge, MA: MIT press.

Ramsey, F. P. (1929). Theories. In Ramsey, FP (1931) *The Foundations of Mathematics and Other Essays* (Braithwaite, R.B., Ed.). London, UK: Routledge and Kegan Paul.

Ramsey, W. (2007). *Representation reconsidered*. Cambridge, UK: Cambridge University Press.

Reich, D. L., Uysal, S., Sliwinski, M., Ergin, M. A., Kahn, R. A., Konstadt, S. N., & Griepp, R. B. (1999). Neuropsychologic outcome after deep hypothermic circulatory arrest in adults. *The Journal of Thoracic and Cardiovascular Surgery*, *117*(1), 156–163. doi: 10.1016/S0022-5223(99)70481-2

Reid, T. (1785/1941). *Essays on the intellectual powers of man*. (Woozley, A. D., Ed.). London, UK: MacMillan and Co, Ltd.

Rescorla, M. (2017). The computational theory of mind. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2017/entries/computational-mind/

Robb, D., & Heil, J. Mental causation In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/entries/mental-causation/

Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition*, *12*(4), 717–731.

Rodriguez-Pereyra, G. (2019). Nominalism in metaphysics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2019/entries/nominalism-metaphysics/

Rosenthal, D., & Weisberg, J. (2008). Higher-order theories of consciousness. *Scholarpedia*, *3*(5), 4407.

Rosemain, M., & Rose, M. (2018). France to spend $1.8 billion on AI to compete with U.S., China. Retrieved from https://www.reuters.com/article/us-france-tech/france-to-spend-1-8-billion-on-ai-to-compete-with-u-s-china-idUSKBN1H51XP

Rovelli, C. (2008). Loop quantum gravity. *Living Reviews in Relativity*, *11*(1), 5. doi: 10.12942/lrr-2008-5

Russell, B. (1919). The philosophy of logical atomism. *The Monist*, *29*(3), 345–380. doi: 10.5840/monist19192922

Ryle, G. (1949). *The concept of mind*. Middlesex, UK: Penguin Books.

Sandberg, A. (2013). Feasibility of whole brain emulation. In V. Müller C. (Ed.), *Philosophy and Theory of Artificial Intelligence. Studies in Applied Philosophy, Epistemology and Rational Ethics* (Vol. 5, pp. 251–264). Berlin, DE: Springer. doi: 10.1007/978-3-642-31674-6_19

Sauchelli, A. (2016). The animal, the corpse, and the remnant person. *Philosophical Studies, 174* (1), 205-218. doi: 10.1007/s11098-016-0677-4

Schaffer, J. (2016). The Metaphysics of causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2016). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/

Scheffler, S. (1982). Ethics, personal identity, and ideals of the person. *Canadian Journal of Philosophy*, *12*(2), 229–246. doi: 10.1080/00455091.1982.10715793

Schneider, S. (2008). Future minds: Transhumanism, cognitive enhancement and the nature of persons. *Neuroethics Publications.* Retrieved from https://repository.upenn.edu/neuroethics_pubs/37

Seager, W. (2007). A brief history of the philosophical problem of consciousness. In Zelazo, P. D., Moscovitch, M., & Thompson, E. (Eds.), *The Cambridge Handbook of Consciousness*. Cambridge, UK: Cambridge University Press.

Sellars, W. (1963). Philosophy and the scientific image of man. *Science, perception and reality, 2*, 35-78. Retrieved from http://www.ditext.com/sellars/psim.html

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. doi: 10.1017/S0140525X00005756

Searle J. R. (1990). Is the brain's mind a computer program? *Scientific America*, (January), 26–31. doi: 10.1038/scientificamerican0190-26

Searle, J. R. (2004). *Mind: A brief introduction*. New York, NY: Oxford University Press.

Selkirk, E. (1995). *Computers for beginners*. New York, NY: Writers and Readers Publishing, Inc.

Serway, R. A., Jewett, J. W., Wilson, .K., Wilson, A, & Rowlands, W. (2018). *Physics for scientists and engineers with modern physics* (2nd ed.). Melbourne, AU: Cengage learning.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Shapiro, L. (Ed.). (2014). *The Routledge handbook of embodied cognition*. New York, NY: Routledge.

Shoemaker, D., & Tobia, K. P. (2019). *Personal identity*. Oxford Handbook of Moral Psychology, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3198090

Shoemaker, S. S. (1959). Personal identity and memory. *The Journal of Philosophy*, *56*(22), 868–882. JSTOR. doi: 10.2307/2022317

Shoemaker, S. (1963). *Self-knowledge and self-identity*. Ithaca, NY: Cornell University Press.

Shoemaker, S. (1970/2008). Persons and their pasts. In Perry, J. (Ed) (Ed.), *Personal Identity* (Second edition, pp. 249–282). Oakland, CA: University of California Press.

Shoemaker, S. (1984). Personal Identity: A materialist account. In S. Shoemaker & R. Swinburne (Co-Authors). *Personal Identity (*pp. 67–132*)*. Oxford, UK: Basil Blackwell.

Shoemaker, S., & Strawson, G. (1999). Self and body. *Aristotelian Society Supplementary Volume*, *73*(1), 287–306. doi: 10.1111/1467-8349.00059

Skinner, B. F. (1953). *Science and human behaviour*. Cambridge, MA: The B.F. Skinner Foundation.

Sloman, A. (1978). *The computer revolution in philosophy: Philosophy science and models of mind*. Sussex, UK. The Harvester Press, Ltd. Retrieved from http://www.csbham.ac.uk/research/projects/cogaff/crp

Sloman, A. (2001). Varieties of affect and the CogAff architecture schema. *Symposium on Emotion, Cognitio, and Affective Computing at the AISB '01 Convention*. Retrieved from http://www.cs.bham.ac.uk/~axs/

Smart, J. J. C. (1959). Sensations and Brain Processes, *Philosophical Review*, *68*, 141-156. doi: 10.2307/2182164

Smart, J. J. C. (2017). The mind/brain identity theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2017/entries/mind-identity/

Snowdon, P. F. (2014). *Persons, animals, ourselves*. Oxford, UK: Oxford University Press.

Staelin, D. H., & Staelin, C. H. (2011). *Models for neural spike computation and cognition*. Seattle, WA: Cognon.net.

Stephan, A. (1999). Varieties of emergentism. *Evolution and Cognition*, *5*(1), 49–59. Retrieved from file:///tmp/mozilla_paul0/Varieties_of_emergentism.pdf

Stoljar, D. (2017). Physicalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/physicalism/

Schwarz, W. (2015). Analytic functionalism. In Loewer, Barry & J. Schaffer (Eds.), *A companion to David Lewis* (pp. 504–518). West Sussex, UK: John Wiley & Sons.

Sugarman, A., & Jaffe, L. S. (1990). Toward a developmental understanding of the self-schema. *Psychoanalysis & Contemporary Thought*, *13*(1), 117–138.

Svensson, H., & Ziemke, T. (2005). Embodied representations: What are the issues? *Proceedings of the Annual Meeting of Cognitive Science Society*, *27*(27), 2116–2121. Retrieved from https://escholarship.org/uc/item/6vm3x1jf

Talbot, M. (n.d.). *The Nature of causation* (No. 1–6). https://podcasts.ox.ac.uk/series/nature-causation

Taliaferro, C. (2018). Substance dualism: A defense. In Loose, Johnathan, J., Menuge, Angus, J. L., & Moreland, J.P. (Eds.), *The Blackwell companion to substance dualism* (pp. 43–60). Hoboken, NJ: John Wiley & Sons.

Thagard, P. (2017). Brain-Mind: From neurons to consciousness and creativity (Draft 5). University of Waterloo.

Thagard, Paul. (2019). Brain–Mind: From neurons to consciousness and creativity. doi:10.1093/oso/9780190678715.001.0001.

The Internet Classics Archive | Theseus by Plutarch (Dryden, John, Trans.). (75 A.C.E). http://classics.mit.edu/Plutarch/theseus.html

Tooley, M. (1990). The nature of causation: A singularist account. *Canadian Journal of Philosophy*, *20*(sup1), 271–322. doi: 10.1080/00455091.1990.10717229

Trappenberg, T. (2009). *Fundamentals of computational neuroscience*. Oxford, UK: Oxford University Press.

Tredoux, C., Foster, D., Allan, A., Cohen, A., & Wassenaar, D. (Eds.). (2005). *Psychology and the Law*. Lansdowne, RSA: JUTA Academic.

Trianni, V., Tuci, E., Passino, K. M., & Marshall, J. A. (2011). Swarm cognition: An interdisciplinary approach to the study of self-organising biological collectives. *Swarm Intelligence*, *5*(1), 3–18. doi: 10.1007/s11721-010-0050-8

Turing, A. M. (1950). Computing machinery and intelligence. In D.C. Dennett & D.R. Hofstader (Eds.), *The mind's I*. New York, NY: Bantam Books.

Tye, M. (2000). Vagueness and reality. *Philosophical Topics*, *28*(1), 195–209. doi: 10.5840/philtopics200028124

Vinge, V. (1993). The coming technological singularity. *Whole Earth Review*, *81*, 88–95. doi: 10.1002/9781118555927.ch35

Wachowski, A., Wachowski, L., Warner Bros., Village Roadshow Pictures, Silver Pictures. (1999). *The matrix*. Burbank, CA: Warner Home Video.

Waish, A. (2017). Saudi Arabia grants citizenship to robot Sophia. Retrieved from https://www.dw.com/en/saudi-arabia-grants-citizenship-to-robot-sophia/a-41150856

Walker, M. (2014). Uploading and personal identity. In R. Blackford & D. Broderick (Eds.), *Intelligence Unbound: The Future of Uploaded and Machine Minds* (pp. 161–177). Hoboken, NJ: Wiley & Blackwell.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*(1), 1–14. doi: 10.1037/h0069608

Welshon, R. (2011). *Philosophy, neuroscience and consciousness*. Stocksfield, UK: Acumen Publishing Ltd.

Wen, H., Han, K., Shi, J., Zhang, Y., Culurciello, E., & Liu, Z. (2018). Deep predictive coding network for object recognition. *ArXiv Preprint ArXiv:1802.04762*. Retrieved from https://arxiv.org/pdf/1802.04762.pdf

Wetzel, L. (2018). Types and tokens. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/fall2018/entriesypes-tokens/

Wiggins, D. (1967). *Identity and spatio-temporal continuity*. Oxford, UK: Basil Blackwell.

Wiley, K. (2014). *A taxonomy and metaphysics of mind-uploading*. Seattle, WA: Humanity+ Press and Alautun Press.

Williams, B. (1970). The self and the future. *Philosophical Review 79*(2), pp. 161–180. doi: 10.2307/2183946

Williams, B. (1973). *Problems of the self*. Cambridge, UK: Cambridge University Press.

Williamson, T., & Stanley, J. (2001). Knowing how. *The Journal of Philosophy*, *98*(8), pp 411–444 Retrieved from http://www.jstor.org/stable/2678403.

Wittgenstein, L. (1953). *Philosophical investigations* (3rd ed.) (Anscombe, G. E. M., Trans.). New York, NY: The Macmillan Company.

Yalowitz, S. (2019). Anomalous Monism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/fall2019/entries/anomalous-monism/