Durham E-Theses

In Defence of the Importance of Substance Ontology for Personal Identity and the Self/Person

GUTA, MIHRETU,PETROS

In Defence of the Importance of Substance Ontology

for Personal Identity and the Self/Person

Mihretu Petros Guta

Doctor of Philosophy

Philosophy Department

Durham University

2014

Abstract

This thesis attempts to show the advantage of substance ontology in providing us the most preferable framework, both on methodological as well as philosophical grounds, to have a better grip on the diachronic problem of personal identity. In this case, substance ontology plays multi-faceted roles in terms of allowing us for example, to make sense of the persistence of persons over time, intrinsic changes persons undergo while maintaining their identity over time, etc. However, substance ontology has not been taken seriously by the majority of philosophers. This is because there is a deep seated but mistaken assumption among contemporary analytic philosophers that given the advances in modern science, substance ontology is irrelevant. It is also not uncommon to see philosophers questionably appealing to an extremely controversial Locke's theory of substratum (i.e., 'something we know not what') to justify their rejection of any notion that goes by the name *substance*. However, with close examination, at the heart of such a rejection of substance ontology lies the naturalistic ontology, according to which everything in the universe has to be explained in purely physical terms as dictated by the physical sciences (e.g., physics and chemistry). But as I will argue in this thesis, when it comes to the metaphysics of the self and its identity over time, the naturalistic ontology suffers from a serious lack of explanatory adequacy. I argue that ultimately, the controversy over the nature of the self is a metaphysical issue in that it is not for science to adjudicate what the nature of the self has to be. In light of this, the conception of substance ontology I defend in this thesis can be taken as Aristotelian in spirit as opposed to Lockean. The category of substance has a fundamental ontological primacy over any other non-substantial entities such as events, places, time, properties (or tropes) and so on. I will argue that substance ontology understood in this way is indeed the most plausible and sustainable conception.

Dedication

This thesis is dedicated to Professor E.J. Lowe, my supervisor, who suddenly passed away while I was wrapping up my thesis. It was a pleasure to study under an extraordinary metaphysician who knew the nuts and bolts of the subject matter. I was not only impressed by his intelligence, but also by his kindness and care for his students. His legacy will continue on.

I also dedicate this thesis to my wife Angela, my son Pete, my mom Aberash Haile and to my sister Nesanet Petros.

Acknowledgements

First of all, I would like to thank my supervisors Professor E.J. Lowe and Dr. Sophie Gibb for their incredible guidance over the past three years.

My thoughts have also been shaped because of the influence of so many other philosophers. I especially want to thank those philosophers with whom I had a chance for face-to-face conversations. In this regard, I thank Saul Kripke (City University of New York), Barry Stroud (University of Berkeley), Lynne Baker-Rudder (University of Massachusetts), Harold Noonan (University of Nottingham), Eric Olson (University of Sheffield), J.P. Moreland (Biola University), Wolfram Hinzen (University of Barcelona/Durham), Robin Hendry (Durham University), Benedict Smith (Durham University), John Heil (St. Louis University Washington), Peter Simons (University of Dublin), Richard King (University of Glasgow), Richard Swinburne (Oxford University) and Jorge Gracia (Buffalo New York State University).

I also had a chance to develop my ideas in various conferences and discussion groups. In this regard, I thank the Ethiopian Graduate School of Theology (EGST) where I served as a guest lecturer on Philosophical Theology in the summer of 2013. While I was at EGST, I had a chance to give a public lecture on part of chapter eight of this thesis. I also thank the Durham Emergence Project Reading Group, where I have had a fascinating exchange of ideas with talented philosophers and scientists. I also thank the Durham Metaphysics Reading Group for excellent weekly meetings and discussions from which I benefited a lot. I also thank the organizers of the Bamberg University International Philosophy Summer School in Germany, which I attended in the summer of 2012. I immensely benefited from my interactions with a group of talented philosophers. I also thank the John Templeton Foundation for inviting me to Ghana to participate in a panel discussion on the ontology of persons in the summer of 2013. I thank the organizers of the Rivendell Institute Summer School at Yale University, which I attended in the summer of 2013. I presented chapter two of this thesis at a philosophy conference organized jointly by the University of Bergen in Norway and Durham University in 2012. I thank the participants from this conference for their comments. I also thank the Durham Eidos Philosophy Community for stimulating discussions over the last three years, where most

Table of Contents

List of Figures

List of Abbreviations and Logical Symbols

    &   Conjunction to be read: 'and'

    V   Disjunction to be read: 'or'

    =   Identity sign to be read: 'the same as' or 'is identical to/with'

    →   If then

    ↔   If and only if

    Iff   If and only if

    ☐   Necessity operator to be read: 'necessarily'

    P   Sentence letter

    TS   Thesis about the Self

    Cat.   Categories

    Met.   Metaphysics

    SM   Serious Metaphysics

    ∀   Universal quantifier to be read: 'for all'

    X   Variable

    Y   Variable

    Rxx   X is related to X

    CSM Cottingham, Stoothoff and Murdoch

Part I: Scene Setting

Chapter One: Introduction

A Survey of the Contemporary Philosophical Debate on the Nature of the Self/Person and of Personal Identity

Before I introduce the thesis outline, in this chapter, I set out to do two main things: First, I will introduce the central question this thesis will attempt to establish and discuss; and second, I will discuss the current state of the debate on the nature of the self and personal identity. I will take up the first task in the first part of this chapter, and the second task in the second part of this chapter. In the first part of this chapter, I will launch into a detailed discussion of the ontology of the self. In this regard, for reasons we shall see, I will focus on Descartes' view of the self. In this thesis, one of my main goals is to argue why we should take the ontology of the self seriously. So from a dialectical point of view, I find it very helpful to begin my discussion with the discussion of the ontology of the self as opposed to mapping out the framework of the contemporary personal identity debate.

In the second half of this chapter, I will locate my discussion of the ontology of the self within the context of the contemporary personal identity debate. In doing so, I will attempt to argue why without taking the ontology of the self seriously, genuine progress is hard to come by in the contemporary personal identity debate. That said, in this thesis, I will make certain assumptions with which everyone will not agree. For example, throughout this thesis, I will argue that genuine self-knowledge or self-awareness is not just a theoretical possibility, but such knowledge is actually attainable. Furthermore, I will also argue that our belief in our continued persistence over time can be given defensible reasons. To make sense of all this, I will argue that our ontology of the self/person must be rooted in our commonsense/folk conception of our own selves. Yet I am also aware of the fact that not everyone is sympathetic to commonsense ontology (see e.g., Churchland 1988). In light of this, the effectiveness of my arguments may be limited. However, I desire that the reader give me the opportunity to show the philosophical work my assumptions accomplish and withhold judgment about the persuasiveness of my thesis until the larger picture has been presented.

1.1. The Question

'Is the Self a Substance?'[1] This question (hereafter *the substantial-self question*) runs together two notions that are hotly debated in the current intellectual climate besetting the issue of personal identity. One is the notion of the 'self'; and the other is the notion of 'substance.' As we shall see, each of these notions has been defined and construed differently by different philosophers. So we might think that the difficulty we face regarding these notions is just dealing with any ambiguities, if such there be, that beset them. Once we dispel any such ambiguities, we might assume that these notions would require no more serious engagement. Unfortunately, that is not the case. This is because, in dealing with these notions, we inevitably come across serious ontological questions lurking right under our nose that need to be taken into account. For example, what is a self? What is a substance? Do such things exist? If so, what is the mode of their existence? In short, are these things real? Contra many contemporary philosophers who are dismissive of such ontological questions (see e.g., Dennett, 1991; Metzinger, 2009; Churchland, 1988), in this thesis I will attempt to show that, in fact, such questions can be given clear and defensible answers.

I will suggest that our inquiry into the problem of personal identity must begin with what I shall call *the Aristotelian question* that takes the form: 'what is X?' Given an inherent distinction between substance and the properties it bears, the category of substance enjoys an ontological primacy over that of the other categories such as qualities. So following Aristotle, I will insist that, in answering the above question, we must make use of 'is' in its primary sense. As Aristotle remarks, "and all other things are said to be because they are, some of them, quantities of that which is in this primary sense, others qualities of it, others affections of it, and others some other determinations of it (*Met*. Z. 1 1028a 18-20; see also *Cat*. 5). The crux of Aristotle's point here is that the existential status of 'beings' belonging to other categories (e.g., qualities, quantities, affections, etc.), depends on the category of

_____

[1] I adopted this question from Ian Gallie's article entitled: 'Is the Self a Substance?' (1936). Gallie himself favors the logical construction view of the self. Grice also (1941; reprinted in Perry, 1975) following Gallie's lead, defends the logical construction theory of the self. To this list, we can add Ayer, who also claims that the self is a logical construction from a series of our experiences (e.g., see 1967: Ch. VII). I will find such views of the self deeply wanting.

substance. This is because non-substantial entities are not self-subsistent and hence they need a bearer, i.e., substance. So in asking the question 'what is X,' we are engaging in a metaphysical inquiry about the nature of X. Hence, I sharply disagree with Ayer, when he says that such questions are "requests for definitions of a particular sort," (1936: 46). At the least, if we follow Aristotle in this matter, our concern is first and foremost to understand the nature of things as opposed to engage in mere linguistic practice (i.e., talk about words).

So we must begin our inquiry into personal identity by asking the question: what is a self/person? As Noonan remarks, "the problem of personal identity is more illuminatingly titled "the problem of the self's identity," (Noonan in Gasser and Stefan 2012: 89). Noonan's remark here reflects two key things to our present discussion. One has to do with the self itself, whereas the other has to do with its (i.e., self's) identity. As we shall see, however, in most contemporary discussions on personal identity the ontology of the self is neglected. Philosophers spend much of their time discussing the identity of the self as if the question of the ontology of the self itself has already been taken care of. In his '*What We Are?*', Olson also remarks, "Why it is that so many philosophers have felt entitled to theorize about personal identity without thinking about what we are, is an interesting question. The answer probably lies in the general neglect of metaphysics throughout the English-speaking philosophy", (2007: 22).[2]

Moreover, I will argue that the issue of the nature of the self and its identity is not only an ontological matter, but it is also a practical matter. As Harold Noonan insightfully remarked:

> What am I? And what is my relationship to the thing I call 'my body'? Thus each of us can pose for himself the philosophical problems of the nature of the self and the relationship between a person and his body. The nature of personal identity over time, and the link, if any, between personal identity and bodily identity are aspects of these problems and it is this, of course, that accounts for the immense

_____

[2] However, once Olson had seriously complained that since there is no agreement on what the notion of the self is, we should entirely stop talking about it (1998). While agreeing with Olson with respect to the lack of agreement among philosophers in characterizing the notion of the self, I find his conclusion to be unjustified for reasons we shall see in this thesis. For various opinions on the notion of the self, see again Gallagher and Shearer, eds., (1999); Gertler and Shapiro, eds., (2007); Gallagher (2000); the self in philosophy, psychology, psychopathology and neuroscience, see Gallagher, ed. (2011).

philosophical interest in the concept of personal identity. But, perhaps unlike some other philosophical problems, the nature of personal identity is not merely of interest to professional philosophers, but also a matter of great practical concern to all of us, philosophers and non-philosophers alike. Man has always hoped to survive his bodily death, and it is a central tenet of many religions that such survival is a reality (1989/2003:1).

One important lesson we can draw from Noonan's remarks here is that the issue of the nature of the self and its identity has two inseparably linked aspects to it: one is theoretical; and the other is practical. On the theoretical side, as we shall see, the focus will be mainly on trying to figure out what the sorts of ontological questions we have raised earlier are supposed to imply regarding us, i.e., human beings. On the practical side, the focus will be on trying to work out what the implications of the answers we give for the ontological questions would mean for us personally. So to have a complete picture of the theory of the self, maintaining the synergy between these two aspects will not be an option. If we could show this to be the case, then all the more it will strengthen Noonan's observations as stated above. But in recent years, as we shall see, of the two aspects I pointed out above, the importance of the practical aspect to the nature of the self has been called into serious question by some influential philosophers like Derek Parfit (1984). In light of this, it could be asked if the proposal for maintaining the synergy between the two aspects will be viable at all. Before addressing such concerns, first we need to locate the above proposal in its proper context. It is only after we do that we can be in a better position to see if there is any justification for it. I will take up this issue beginning in § 1.2.

That said, returning to *the substantial-self question*, one of the things we are invited to do is to affirm or to deny whether or not the self belongs to the category of substance. To clearly see how this is supposed to work, first we need to turn *the substantial-self question* into an evaluable statement that takes the form: '*X is F*'. Here we can substitute the variable '*X*' for 'the self' and the predicate '*F*' for 'a substance'. In doing so, we can construct a statement as follows: '*the self is a substance*.' Although put in this way, the statement makes perfect grammatical sense from the standpoint of the present discussion, however, it faces an underlying semantic ambiguity. To show the nature of such an ambiguity, I adopt the

distinctions E.J. Lowe (1989:3-4) draws on the varieties of the copula (i.e., 'is'). Here they are:

> (a) The 'is' of predication/*attribution* (e.g., 'Plato *is* intelligent', 'Sky *is* blue').
> (b) The 'is' of identity (e.g., 'Water *is* H_2O').
> (c) The 'is' of instantiation (e.g., 'A kangaroo *is* a mammal').
> (d) The 'is' of constitution (e.g., 'A human brain *is* a collection of cells').[3]

Engaging in full-fledged analysis of (a)-(d), would take us too far afield. So I only want to base my discussion on this particular question: Which reading of 'is' should we adopt to grasp the semantics of 'is' in the statement, '*the self is a substance*'? One way we can answer this question is by applying each reading of 'is' in (a)-(d) to that of the above statement and see which reading captures its meaning. There are different ways one could take in dealing with this issue. Here I go along with the Aristotelian conception of substance (see *Cat*. 2ª 11ff).[4] Since Aristotle's view of substance will be discussed throughout this thesis, I will be brief in what follows.

According to Aristotle, *inter alia*, a substance is a property bearer. Being a property bearer, a substance is a subject of predication, i.e. other things are predicated of it but not vice versa. (e.g., 'Socrates is white'). Aristotle gave a special term for subjects of predication—*primary substances*, namely 'the individual man' (*Cat*. 2ª11). In light of this, applying (a)'s reading to the above statement could be seriously misleading, if we understand 'a substance' to mean a primary substance. This is because, primary members of the category of substance, i.e. 'individual Socrates' are not predicates but subjects (see further Ross, 1924:1xxxii-xc).[5] Hence, (a)'s reading in relation to the above statement (i.e., the self is a

---

[4] But defenders of bundle and substratum theories of substance could take an approach that is opposed to the one I am adopting here (see e.g., Loux, 1998: Ch.3).

[5] However, on Aristotle's view, there is a sense in which a substance can be said a predicate. Aristotle assigns for this notion another special term—secondary substances, namely 'man' and 'animal' (*Ibid*.). Aristotle calls secondary substances kind universals (e.g. humanity) as opposed to mere property (e.g., blueness). In illustrating this latter point, Ross remarks, "What is this thing? A man. What is a man? An Animal. What is an animal? A substance.' 'Substance' is the last predicate we come to if we pursue such a line of inquiry…" (Ross, 1924: 1 xxxiv). Although for now we are not interested in exploring the relation of this latter notion of substance to the above statement (i.e., the self is a substance), we will return to it in due time. As we shall see, the two notions of substance are interdependent in an ontological sense of the term, i.e., one needs the other for its existence.

a substance'), remains to be problematic. How about the (b) reading? By applying this reading to the above statement, we are committed to taking the terms 'the self' on the one hand, and 'a substance' on the other hand, as co-referring expressions. That means that these expressions must have the same referent. So far so good. But the problem we encounter with (b)'s reading comes when we begin to give an account of the nature/identity of this referent to which both of the above expressions are meant to stand for. This is the point of serious contention in philosophy of mind, between the proponents of substance dualism (one brand or another) and physicalism (one brand or another). While the former group, accepts the duality of substance (i.e., nonphysical and physical), the latter group endorses only a non-dual conception of substance (i.e., physical). I will take up this issue in chapter six. For now I content myself with the above brief analysis. But as it stands, the (b) reading does not pose any serious problem to the above statement.

That said, what are we to make of the remaining two readings, i.e., (c) and (d) in relation to the above statement? Given the Aristotelian framework introduced in the *Categories,* applying the (c) reading to the above statement will be problematic. This is because *primary substances*, although they can be instances of, for example, a kind universal—man or humanity, they themselves cannot be instantiated by any other entity (cf. footnote # 4; also see Loux, 1978: 158-163; Lowe, 2000: 350-353). So applying the (c) reading to the above statement simply misplaces the *primary substance* in the domain of that of a *secondary substance*, the very point that is being disputed here. So the prospect for the (c) reading in helping us capture the semantics of 'is' in the above statement is no better than that of the (a) reading. Finally, the (d) reading itself does not seem to fair any better either. To see this, I want to bring out a further analysis of the notion of substance, as I shall call it, in terms of in its *scientific sense* on the one hand, and in its *metaphysical sense* on the other. Here we can capture the scientific sense of substance, via philosophers' favorite example 'the statue and the bronze.' According to one view (which I am inclined to endorse), the relation between the statue and the bronze is not based on identity (see Aristotle *Met. Z.* 1033a 4-20, Z. 8 1033b 1-9; Baker, 2000: Ch.2). That is, the bronze is what constitutes the statue of which it is made. In other words, the statue, say of Goliath, is distinct from the bronze (i.e., staff) that constitutes it. Moreover, Goliath and the lump of bronze, *inter alia*, have

18

different modal properties. For example, Goliath can't survive if it is squashed and turned into another statue, say of a watch. But the lump of bronze survives such a change. On the other hand, Goliath survives the loss of some of its part (e.g., a finger). But the same thing is not true of the lump of bronze, since any loss of its particles would compromise its identity. At least, at this point, I do not wish to enter into a contemporary debate over the relationship between the statue and the bronze. My aim here is only to illustrate the point that, taken in its scientific sense, *stuff* or aggregates of matter can be said to be a substance. Such talk is in fact prevalent in chemistry, where various chemicals/elements are often described in the language of substance (see Connell, 1988: 10).

But having said that, if we still stay within the Aristotle's *Categories*, the scientific sense of substance plays no role whatsoever in terms of helping us capture the meaning of 'is' in the above statement (i.e., '*the self is a substance*'). This is because, in the *Categories*, Aristotle is working with the *metaphysical* notion of substance for which, he gives a technical Greek word, ουσια/*ousia*, the noun from the verb 'to be' (*einai*). For Aristotle, *ousia* means, among other things, fundamental existent (see *Cat*. 2ª11). Although the English term 'substance' corresponds to the Greek word *ousia*, the former is a conventional rendering, which came from the Latin substantia, i.e., 'something stands under.' However, both 'substance' and substantia fail to capture the literal meaning of *ousia*. In fact, the English term 'substance' could be misleading, implying a kind of stuff (Cohen 2009: 197). The closest literal translation that does justice to *ousia* is 'being' or 'entity' (Ackrill 1963: 77). *Ousia* can also be construed as 'reality' or 'fundamental being' (Cohen 2009: 197) or 'an ontologically basic entity' (Loux 1991: 2 ff). But things are not that straightforward when it comes to Aristotle's *Metaphysics*, where the notions of 'matter' and 'form' are introduced (see e.g., *Met.* Z-H). Even then, as I will argue, Aristotle did not detract from his original position, where he takes *primary substances* (*ousia*) as fundamental entities, i.e., other things depend on them but the converse does not hold. In light of such considerations, therefore, there seems to be no plausible way by which we can apply the (d) reading to the above statement.

Although as we have seen, the alternatives (a)-(d) all have their drawbacks, from the standpoint of the contemporary discussions on the nature of the self, the (b) reading still has central importance. Given the (b) reading, therefore, if one affirms that the self belongs to the category of substance, then whatever characterization we give to the notion of 'substance' will apply to the 'self.' In this case, the converse would also be true. On the contrary, if one denies that the self belongs to the category of substance, then the link between the notion of the 'self' and the notion of 'substance' would collapse. That means that we cannot apply the characterizations we give to the notion of 'substance' to that of the self and vice versa. Currently the prevailing assumption among the contemporary philosophers is that the era of understanding the self as a substance has gone out of favor. Consequently, for most, if not all analytic philosophers, *the substantial-self question* has lost its relevance.

For example, Sydney Shoemaker once remarked that nowadays the question of whether the self is a substance and whether personal identity over time requires a substance, has a musty smell to it. Despite the fact that the question in the past played a central role in discussions of personal identity in Locke, Butler, Hume, and Reid, it has not been the central question in contemporary discussions of personal identity. In most such discussions, the question simply is not addressed (2003: 381).[6] Here I take Shoemaker's observation as indicative of three distinct, but interrelated issues that need to be examined carefully. These issues are: (i) the rejection of *the substantial-self* for personal identity over time; (ii) the rejection of the centrality of *the substantial-self question* for personal identity; and (iii) the absence of *the substantial-self* question from most discussions of personal identity, respectively.[7] Here (i)-(iii) together constitute what I shall call: *the marginalization of the ontology of the self*. But as will be shown throughout this thesis, the marginalization of the

_____
[6] But in his *Self-Knowledge and Self-Identity*, Shoemaker claims that philosophical discussions of the nature of persons/selves have often focused on the question, 'Is the self a substance?' (1963: 41). In section 2.1, we will see the reason why in recent years this question has been pushed aside by most personal identity theorists.
[7] A detailed discussion of these issues will be given in chapters two through eight.

ontology of the self in contemporary analytic philosophy is primarily rooted in (i). But at this point, the question that comes to mind is this: why do most analytic philosophers unite in embracing (i)? To answer this question, first we need to answer our original question, i.e., *the substantial-self question*. Here I will begin my discussion with a well-known classical response to *the substantial-self question*.

1.1.1.   Rene Descartes's Answer:

Descartes's answer to *the substantial-self question* was responsible in setting the initial tone for the modern philosophical debate on the nature of the self (see e.g., Frondizi, 1953: Ch. 1; Armstrong 1999: 2; Atkins 2005: 7). Moreover, as we shall see, Descartes's answer to *the substantial-self question* provided a backdrop against which subsequent philosophers (e.g. Locke) developed their own alternative views of the self. But before we turn to Descartes's answer, it is worth reminding ourselves of the remark that was made in § 1.1, with respect to what we are invited to do with respect to *the substantial-self question*. We had said that this question invites us to affirm or deny whether or not the self belongs to the category of substance. If the self belongs to the category of substance then whatever characterization we give to the one equally applies to the other. That means that both terms, i.e., the self on the one hand and a substance on the other hand, refer to one and the same thing (i.e., entity). But how are we supposed to show this? This is where Descartes's answer to *the substantial-self question* comes into play.

Descartes's answer emerges from two standpoints—epistemological as well as ontological. Taken from the standpoint of epistemology (i.e., theory of knowledge), Descartes's focus is on the question of whether or not we know that the self exists. On the other hand, taken from the standpoint of ontology (i.e., theory of the nature of being), Descartes's emphasis is on spelling out the nature of the self, i.e., what kind of thing that it is. Let us now look briefly at each one of these points.

1.1.1.1 The Epistemology of the Self

Descartes develops his epistemology of the self within a framework of his general theory of knowledge. Central to Descartes's general theory of knowledge is the notion that genuine

21

knowledge must be free from any error (see *Meditations on First Philosophy* AT VII: CSM 1984; also see Bonjour 2010: Ch. 2; Feldman 2003: 52-60). But how did Descartes go about ensuring error free knowledge? Details aside, Descartes's uses his well-known methodic *doubt* to filter out instances of genuine knowledge from those that are not (cf. Williams, ed., 1993: Introduction; Stroud 1984: Ch. 1). For example, Descartes rejects the senses as being genuine sources of knowledge arguing that in principle they could be susceptible to serious deception (e.g., evil demons, dream states), thereby producing false beliefs (see *First Meditations* AT VII 18-23: CSM 12-15). That said, when it comes to the epistemology of the self, the only *genuine knowledge producing source*, as Descartes sees it, is *a priori* introspective reflection—rooted in the first person perspective (see *Second Meditation* AT VII 24-34:CSM 16-23). It is in this sense that Descartes speaks of self-knowledge. As Descartes remarks:

> While I was trying thus to think everything false, it was necessary that I, who was thinking this, was something. And observing that this '*I am thinking, therefore exist*' [*Cogito ergo sum*] was so firm and sure that all the most extravagant suppositions of the sceptics were incapable of shaking it, I decided that I could accept it without scruple as the first principle of the philosophy I was seeking (*Discourse* AT VI 32: CSM I 127).

Descartes's point here is that an act of thinking is inextricably linked with a thinker's existence. So given Descartes view, the sense of the phrase 'inextricably linked' needs to be understood both from the epistemic as well as the ontological standpoints. It is epistemic in that an act of thinking provides one with the basis for one's self-knowledge. It is also ontological in that thinking strongly presupposes a thinking being. This is the central point that underlies Descartes's famous Latin dictum, *Cogito ergo sum* (*Cogito* for short). *Cogito* terminated Descartes doubt by assuring him of his existence (cf. Frege 1956).[8] Long before Descartes, Augustine also had invoked the *Cogito* argument to combat the skepticism of the *Academy* (see e.g., *De civitate dei*, XIII; XI: 26; *De trinitate*, Bk. X. Ch. 10; also cf. Ganeri, 2012: 57-60; Menn, 1998; Kehr, 1916). But it was Descartes who appropriated the *Cogito* explicitly in the first person stance (see Anscombe in Cassam 1994: 140). In doing so,

---

[8] Descartes is not sceptic in an ancient Pyrronian sense, according to which knowledge is not possible. Descartes's skepticism is merely methodological—aimed at establishing a secure foundation for knowledge.

Descartes shows why one cannot succeed in doubting one's own existence, without thereby implying a prior grasp of one's existence. As Thomas Nagel nicely sums up:

> In the *Cogito* the reliance on reason is made, explicit, revealing a limit to this type of doubt. The true philosophical point consists not in Descartes' conclusion that he exists….nor even in the discovery of something absolutely certain. Rather, the point is that Descartes reveals that there are some thoughts which we cannot get outside of (Nagel 1997: 19; also cf. Frege 1956; Mctaggart Vol. 1 1921: 56-58).

So we can sum up Descartes's official view of the epistemology of the self as follows: the self knows its existence, because its existence is firmly secured in first person awareness of itself. That said, notice that Descartes's *Cogito* does not specify the nature of this self. This is a deliberate move on Descartes's part, since here his focus is on showing the role of the first person thought in establishing one of the key aspects of self-knowledge, i.e., one's knowledge of one's own particular thoughts (cf. Cassam 1994: 1). At least for now, the "I", in the *Cogito* functions in what I shall call*: a neutral fashion.* Here by *'a neutral fashion',* I mean that the "I" in *the Cogito* does not necessarily favor, for example, the substance dualists' understanding of it (e.g., immaterial thing) over that of the materialists'/physicalists' (e.g., material thing) and vice versa. Hence, one need not subscribe himself/herself to either position to be able to see the gist of Descartes's *Cogito*.

Here an example from Ludwig's Wittgenstein's *The Blue and Brown Books* will bring my point home. In the *Blue Book*, Wittgenstein distinguishes between two uses of "I": (a) "the use as object"; and (b) "the use as subject" (1958: 66). An example of (a) is "my arm is broken" and an example of (b) is "I think it will rain", "I have a toothache" (*Ibid*. 66-67). Details aside for now, for Wittgenstein, the category of (a) admits an error. For example, as Wittgenstein points out, in an accident it is possible that one may feel a pain in one's arm, if one sees a broken arm at one's side thinking that it is one's own arm (i.e., thereby forming a false belief). But in the case of the category of (b) there is no possibility for an error. For example, if I am in pain, then I cannot be mistaken about who is in pain. This is because my awareness of my own pain experience is squarely rooted in my first person awareness of it.

So the semantics of "I" in the *Cogito* aligns with Wittgenstein's (b) as opposed to (a).[9] Taken this way, it seems to me that primarily the mode of self-knowledge implied in Descartes's *Cogito*, is not propositional in nature, which takes the form: *I know that P*. Rather, it is what Bertrand Russell of *The Problems of Philosophy* describes as: knowledge by acquaintance, which takes the form: *I know P*.[10] So we can be said to know something in the sense of knowledge by acquaintance, as Russell remarks, "[when] we have acquaintance with anything of which we are directly aware, without the intermediary of any process of inference," (1959: 25). Here Russell's point is that the things that we come to know by direct acquaintance are not the result of, for example, a certain conclusion derived from a set of premises. It is quite the opposite. That is, as I understand it, *knowledge by acquaintance* is an instance of knowledge, whereby the subject finds himself/herself simply knowing something. Here by 'simply knowing something', I mean with no precondition put in place on the basis of which one comes to know something. Instead the locution, 'simply knowing something' stands for, as it were, an immediate grasp of something. For example, if I am in a pain state, then I know the state I am in directly. If I am asked how I know that I am in a pain state, the right thing for me to say is this: *I just know it*. Hence, in such contexts, seeking an inference based answer for the '*how*' question effectively becomes meaningless or so it seems to me. This is because my knowledge of my being in a pain state is not dependent on me first doing something else before I come to know it. Gareth Evans seems to be thinking along similar lines when he says:

> A subject's knowledge of what it is for the thought 'I am in pain' to be true may appear to be exhausted by his capacity to decide, simply upon the basis of how he feels, whether or not it is true—and similarly in the case of all the other ways of gaining knowledge about ourselves (Evans in Cassam 1994: 187).

In light of this, if my brief analysis above on *Cogito* is right, then we can say that the knowledge implied by "I" in Descartes's *Cogito* is ultimately *de re* (i.e., a thing) knowledge as opposed to being merely *de dicto* (i.e., propositional) knowledge (cf. Chisholm 1976).[11] That

---

[9] Dorit Bar-On also advances similar arguments in her excellent book, *Speaking My Mind: Expression and Self-Knowledge* (e.g., see 2004: Ch. 2).

[10] However, Russell's own view on this issue underwent change. But that does not affect my point here.

[11] For an excellent discussion concerning *De re* and *De dicto* modality, see Plantinga (1974: 1-36); also see Davidson, ed., (2003: Ch. 1).

is, what Descartes claims to have known via *Cogito* is quite literally *himself* (cf. Merleau-Ponty, Part Three, 1962: Ch. 1). Put differently, for Descartes, *Cogito* provides an epistemic basis for *de re* knowledge. This seems to be the background assumption for Descartes when he says in his *Second Meditation*: "this proposition, I am, I exist, necessarily true whenever it is put forward by me or conceived in my mind," (AT VII 25: CSM 17). In light of this, if my reading of *Cogito* into Russell's conception of knowledge by acquaintance holds true, then we can plausibly assume that ultimately one's knowledge of oneself is not theory laden at all, contra to current orthodox accounts in analytic philosophy in general and cognitive neuroscience in particular. This is the view I very much want to defend in this thesis. I shall call this: *a non-theory laden conception of the self/person*. I will pick up this discussion in chapter seven.

That said, before I end my discussion in this section, I anticipate one objection (to be discussed in chapter eight) against my hitherto analysis of *Cogito*. For example, it may be said that both Descartes's *Cogito* in general and the analysis I have given for it in particular suffer from circularity. For example, in his *The Paradox of Self-Consciousness*, José Luis Bermúdez gives the most sophisticated account of such circularity. But I should note that Bermúdez's own account is not given in consideration of Descartes's *Cogito*. Regardless of that, however, my interest in Bermúdez's objection has to do with its direct relevance to the present discussion, as well as to what I briefly indicated earlier as a non-theory laden conception of the self/person. With this caveat in mind, here are the two types of circularity that Bermúdez discusses:

> Any theory that tries to elucidate the capacity to think first-person thoughts through linguistic mastery of the first-person pronoun will be circular, because the explanandum is part of the explanans either directly…or indirectly. Let me call this *explanatory circularity* (1998: 16).

Bermúdez also identifies what he calls *capacity circularity*:

> The capacity for reflexive self-reference by means of the first-person pronoun presupposes the capacity to think thoughts with first-person contents, and hence cannot be deployed to explain that capacity. In other words, a degree of self-consciousness is required to master the use of the first-person pronoun (1998: 18).

Bermúdez claims that both the explanatory circularity and the capacity circularity constitute what he calls: *the paradox of self-consciousness* (see 1998: 14-24). Bermúdez also suggests the way out of both forms of circularity. He sums it up as follows:

> A nonconceptual content is one that can be ascribed to a thinker even though that thinker does not possess the concepts required to specify that content. Nonconceptual first-person contents are those that fall into this category and that can be specified by means of the first-person or indirect reflexive pronouns. This nonconceptual first-person content offers a way of breaking both forms of circularity (1998: 49ff).

We have yet to see whether the circularity Bermúdez pointed out above can be successfully avoided. Also I remain unpersuaded by Bermúdez's ingenious solution aimed at avoiding such circularity. I will argue that Bermúdez's account of circularity suffers from the conflation of a disposition with its manifestation. Contra Bermúdez, as I will argue, if the nonconceptual first-person content is understood in dispositional terms, then under the right circumstance, its manifestation will take place in terms of a capacity, namely for reflexive self-reference via the use of the first person pronoun. So once we take a dispositional approach to understand nonconceptual first person content, Bermúdez's explanatory as well as capacity circularity turn out to be harmless.

The focus of our hitherto discussion has been on the first aspect of Descartes's answer to *the substantial-self question*, i.e., the epistemic aspect of the existence of the self. We now turn to the second aspect of Descartes's answer to *the substantial-self question*—which deals with the question: what kind of a thing the self is.

1.1.1.2 The Ontology of the Self

Descartes develops his ontology of the self within a framework of his metaphysics of substance*.* As we recall in § 1.1.1, we had said that in the context of the modern discussion of the ontology of the self, Descartes's answer to *the substantial-self question* gives us the first instance of the identification of the self with a substance. Here is the first clear example of such identification:

> I knew I was a substance whose whole essence or nature is simply to think, and which does not require any place, or depend on any material thing, in order to exist. Accordingly this 'I'-that is, the soul by which I am what I am-is entirely distinct from

the body, and indeed is easier to know than the body, and would not fail to be whatever it is, even if the body did not exist (*Discourse* AT VI 33 CSM: 127).

This passage encapsulates three key aspects of Descartes's characterization of the nature of the I, i.e., the self; (i) I is a substance, whose nature wholly consists in thinking; (ii) "I" is a referring expression, which in this case stands for a thinking thing; and (iii) I as a substantial thinking thing is: nonspatial, independent of and distinct from the material thing (i.e., the body), is capable of existing without the body and maintains self-identity over time. Of (i)-(iii), as we shall see, the most controversial aspect of Descartes's characterization of I remains to be (iii). It is also (iii) that underlies Descartes's well known: *the mind-body problem*. But what is the metaphysical basis of (iii)?[12] This is the question that Descartes deals with in great detail both in his *Second* and *Sixth Meditations*. But for present purposes, I do not concern myself with these details since my goal here is only to look at the second aspect of Descartes's answer to *the substantial-self question* exclusively from the standpoint of substance ontology.[13]

In this regard, both in *Second* and *Sixth Meditations*, Descartes expands upon (i)-(iii), where his main focus is on establishing the ontological status of the self.[14] For example, Descartes

_____

[12] For details on Descartes's mind-body problem (e.g., the problem of psychophysical causal interaction, see Kim (2006; 2005: Ch. 3); Heil (1998); Cottingham (1986: Ch. 5). In chapter six, I will discuss some issues in philosophy of mind in relation to the self.

[13] In *Principles of Philosophy* Descartes speaks of two notions of substance in light of which he laid the initial ground for his discussion in *Meditations*. First, a substance is a thing that exists independently, i.e., exists by itself. As Descartes states, "By substance we can understand nothing other than a thing which exists in such a way as to depend on no other thing for its existence," (*Principles* AT VIII I.51 CSM: 210). Cottingham (1993: 160) remarks that here, Descartes's definition of substance directly echoes Aristotle's view of substance, according to which "a substance is that which is neither said of a subject nor in a subject, e.g., the individual man or the individual horse," (*Cat*. 2a 11). As Cottingham further points out, "the contrast here is between things which exist independently (such as individual horse) and properties or attributes (such as being fleet-footed) which can only be predicated of, or belong to, a subject," (*Ibid*.). Second, a substance is a bearer of properties. For example, a material substance or the physical body bears properties such as extension in space, size, figure whereas the immaterial substance is a bearer of thought (*Principles* AT VIII I. 48-53 CSM: 210). That said, for Descartes, only God satisfies the requirement of independent existence in a sense of not needing to depend on anything whatsoever. Created things need God's concurrence in order to exist (see for details, Woolhouse 1993: Ch. 2). So the notion of substance cannot be used univocally, i.e., in the same sense for God and other created entities (see, *Principles* I.51 CSM: 210). In his own way, Spinoza also defends an Aristotelian notion of the independent existence of substance, see e.g., *Ethics* trans., Parkinson (2000: 15-17). For sophisticated account of what an independent existence amounts to, see Lowe (2009).

[14] Descartes uses terms such as: mind, self and soul and "I" interchangeably.

asks, "But what then am I?" (*Second Meditation* AT VII 28 CSM: 19). He answers this question by saying that he is, "a thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions," (*Ibid*.). Here Descartes's answer runs together three key features that constitute the nature of the self, which will also be the focus of my attention throughout this thesis. The features I have in mind are: the self as an agent (e.g., willing and unwilling); the self as a bearer of cognitive/mental states (e.g., doubt, affirms); and the self as a distinct entity—from the properties it bears. In his *Sixth Meditation*, Descartes expands upon these features mainly by drawing the distinction between the body and the mind on the one hand, and the role they each play on the other hand (see AT VII 72-90 CSM: 50-62). For example, Descartes argues that only an incorporeal substance is the subject of mental properties. Likewise, only corporeal substance (i.e., physical body) is a bearer of physical properties such as motion, shape, figure, size, etc., (cf. footnote # 12 above). In light of such considerations, Descartes gave an affirmative answer to *the substantial-self question* i.e., the self (i.e., immaterial) is a substance.

1.1.2 Alternative answers

As indicated in § 1.1.1, I pointed out that Descartes's answer to *the substantial-self question* was responsible for subsequent alternative philosophical theories of the self. For present purposes, there is no need to list all of these alternative theories of the self nor will it be practical to discuss them on an individual basis.[15] So, in this section, I will suggest five representative categories under which alternative answers to the *substantial-self question* can be grouped. These are: (a) broadly Cartesian answers; (b) broadly Lockean answers; (c) broadly Humean answers; (d) broadly Kantian answers and (e) broadly Aristotelian answers.[16] However, in suggesting these categories, I am not implying, for example, that

---

[15] Outside of analytic philosophy, the notion of the 'self' crops up in many other places as well (e.g., in sociology, in cultural anthropology, in psychology, in political theory, in world religions, in films, fictional/literary writings, just to mention a few). In this thesis, however, my discussion of the notion of the self will be confined entirely to analytic philosophy. That said, for a nice representative view of the self both from analytic and continental philosophical traditions, see Atkins (2005). Specifically, for the theory of the self in the transcendental tradition, see Carr (1999).

[16] Here the ideas in one way or another are linked to Descartes, Locke, Hume, Kant and Aristotle.

any given answer to the *substantial-self question* neatly fits only into one of the above domains/categories. On the contrary, as we shall see, answers to *the substantial-self question* may well be partly Lockean in one sense and partly Humean in another sense and so on. So we should not necessarily take these categories as rigid benchmarks against which we determine which category any given answer to *the substantial-self question* must belong.

1.1.2.1 The Two Reasons

That said, at this point, the question that comes to mind is this: why did philosophers suggest alternative answers to *the substantial-self question*? In other words, why did they refuse to settle for Descartes's answer? While a full-scale answer to this question will take us too far afield, in the context of the present discussion, two reasons come to mind: (1) Descartes's *identification of the self with a substance*; and (2) Descartes's *recognition of the self as a* metaphysical entity. Here (1) and (2) are inextricably interrelated. In (1) the term 'identification,' should be understood in light of our discussion in § 1.1 and § 1.1.1.2. In these sections, we saw that the terms the 'self' and a 'substance' refer to the same thing—thereby implying strict logical identity. On the other hand, in (2) the term 'recognition,' should be understood from an epistemic standpoint in that it is intended to show that one thing is being recognized as something. Such knowledge, for example, takes the form: *I know P as Q.* As we recall, Descartes speaks of knowing himself *as a thing* whose essence is to think (see § 1.1.1.2). In light of this, I shall call the 'identification,' in (1) *Metaphysical Identification*; and the 'recognition,' in (2) *Epistemological Recognition*. Having clarified the sense in which we are using the terms 'identification' and 'recognition' we are now in a position to see why the majority of analytic philosophers distance themselves from embracing Descartes's answer to *the substantial-self question*.

1.1.2.2 Locating the Objection

One way we can show this is by looking at Descartes's substance ontology, which, in light of our discussion in § 1.1.1.2 can be summed up under the following three definitions:

(D1) *X* is a substance=$_{df}$ (a) *X* exists; (b) *X* does not depend on *Y* for its existence but *Y* depends on *X*. (*Y* refers to creatures such as human beings).

(D2) *X* is an incorporeal substance=$_{df}$ (a) *X* exists; (b) *X* is a non-physical thing: (i) *X* does not occupy space, (ii) *X* does not depend on any material thing, and (iii) *X* is capable of existing without the body, (iv) *X* maintains strict identity over time; (c) *X* is a bearer of mental properties *M*; (d) mental properties *M* depend on *X* for their existence but *X* does not depend on *M* for its existence.

(D3) *Y* is a corporeal substance=$_{df}$ (a) *Y* exists; (b) *Y* is a physical thing; (c) *Y* is a bearer of physical properties *P*; (d) physical properties *P* depend on *Y* and *Y* also depends on *P*.

The majority of contemporary analytic philosophers who are critical of Descartes's view of the self exclusively target (D2).[17] The variable '*X*' in the *definiens* of (D2) is to be substituted for the term 'self'. To clearly see the philosophers' reasons for rejecting (D2), it would be helpful to analyze the objection raised against it at three levels. Doing this is extremely important, since objections against Descartes's (D2) often go unchecked. So we need to show which objections are effective and which ones are not. I shall call the three levels as follows: the *existence,* the *terminology* and the *unity levels*, respectively.

A*. The Existence Level*

At the *existence level*, the objection mainly consists of the conjunction of *D2*-(a) and *D2*-(b)'s main *definiens* i.e. the self is a non-physical thing. The nature of this objection can be stated as follows: since the self is said to be a non-physical thing and a non-physical thing is unobservable, it follows that its existence cannot be established (see e.g., Quinton in Perry, ed.: 1975: 54-55; Perry 1977; Glover 1988: 98; Dennett 1991; Kolak and Martin, eds., 1991: 339). But we have yet to see whether this conclusion follows. That said, it is not difficult to pin down the assumption behind this objection. That is, if *X* exists, then *X* must be an object of observation. But what if *X* is not observable? If we take the above assumption at face value, then what we should say in response to this question seems to be this: if X is not

---

[17] In Descartes's time, Hobbes rejected (D2), see *Leviathan*, ed. (1994); see also Gassendi's *First Set of Objections*, AT VII 257-277: CSM 179-193; Descartes's Reply: AT VII 351-361: CSM 243-249; AT VII 385-391: CSM 264-267; also see *Seventh Set of Objections* with *the Author's Replies*, AT VII 451-561: CSM 302-383. For early Wittgenstein's (i.e., *Tractatus*) objection against (D2) see Kripke's excellent discussion (1982: 122-145). All of these objections can be subsumed under what I called objection at *the existence level* (see A below).

observable, then it does not exist.[18] Such a response in turn leads us to another question: Does unobservability necessarily imply non-existence? More to the point: Is the self's unobservability (whatever that means) necessarily imply its non-existence? Here I inserted a qualifying phrase 'whatever that means,' so as to show in the course of our discussion that, the very notion of 'the observability of the self' is deeply misguided—which needs to be abandoned altogether. I will return to this issue in chapter three. But before I answer the above question, I need to mention two key historical precedents that are directly responsible for the objection under consideration. The first one has to do with the advent of modern science; and the second one has to do with Hume's extremely influential answer. For example, regarding the former, Dennett remarks:

> Since the dawn of modern science in the seventeenth century, there has been nearly unanimous agreement that the self, whatever it is, would be invisible under a microscope, and invisible to introspection, too (1991: 412).

With respect to the latter, in his *A Treatise of Human Nature*, Hume remarks:

> For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can observe any thing but the perception….If any one upon serious and unprejudic'd reflexion, thinks he has a different notion of *himself,* I must confess I can reason no longer with him….tho' I am certain there is no such principle in me (I. IV. VI 1888: 252).

In his remarks, Dennett has already implicitly alluded to Hume's view, as the phrase 'invisible to introspection' indicates.[19] So in this case, Dennett's remarks nicely encapsulate the unobservability of the self both from the standpoint of scientific experiment on the one hand and one's introspection on the other. At this point, many philosophers (including

---

[18] Of course, here one can ask: what is existence? Or alternatively: what does it mean for something to exist? As is well-known, in the post Frege-Russell-Quine era, the debate besetting the notion of existence focuses, inter alia, on the analysis of existential quantification and reference (see .e.g., Aune 1985: Ch. 2; Williams 1981; Quine 1969: Ch. 4, Russell in Marsh 1956). For present purposes, I do not need to get into this highly technical area of logic and philosophy of language. In *The Search for Truth*, Descartes states: "To know what that is, all we need do is to understand the meaning of the word [i.e., existence], for that tells us at once what the thing is which the word stands for, in so far as we can know it. There is no need for definition, which would confuse rather than clarify the issue," (AT X 525: CSM II 418). But Descartes's proposal here need not be seen along the lines of Frege's, Russell's and Quine's analysis of the notion of existence. As we already saw in the *Cogito* case, Descartes seems to be working with an intuitive notion of existence, which in the words of Mctaggart, 'a species of the real' (see e.g., Mctaggart Vol. I 2 1921: 1-2, 57-58).

[19] I will discuss Hume's view in detail in chapter three. So I will not engage in serious discussion yet.

Dennett), assume that Descartes's D2-(b)'s main *definiens* (i.e., the non-physicality of the self) justifies their denial of D2-(a), i.e., the self's existence. In the next section, I will question the plausibility of such a move. More importantly, I will point out that those who reject Descartes's D2-(a) on the basis of D2-(b)'s main *definiens* are trading on an ambiguous term 'observation.' Before I discuss that, I now turn to the terminology level since it is directly linked to the existence level.

B. *The Terminology Level*

As we saw at *the existence level*, the main objection that was raised against the existence of the self has to do with the alleged undetectability or unobservability. Those who embrace the 'undetectability of the self' objection also raise other objections at the terminological level. In this regard, the objections stem from the four terms Descartes interchangeably uses to refer to himself. These terms are: 'mind', 'soul', 'I' and 'self' (see footnote # 12 in § 1.1.1.2). As summed up in (D2) in §1.1.2.2, for Descartes, all of these terms refer to a non-physical entity/thing. But since the 'undetectability objection' rules out the existence of a non-physical entity, Descartes's opponents conclude that the above four terms do not have a referent. So to think otherwise, according to Descartes's opponents, would be to endorse what Gilbert Ryle (objecting to Descartes's view of the self) describes as, "the dogma of the Ghost in the Machine," (1949: 15). Furthermore, Ryle claims that Descartes's view of the self is a category mistake, which is false not just in detail but also *in principle* (emphasis mine, *Ibid*.16).

But Ryle's remarks above are not only questionable but are also extremely misleading. If as Ryle boldly asserts that Descartes's view is false *in principle*, then such remarks do seem to suggest that there is nothing we can positively learn from what Descartes has to say about the ontology of the self. Contra Ryle, therefore, I will argue that there are things we can positively learn from Descartes's view of the self. As they stand, Ryle's remarks above can only be taken as a paradigm example of an extreme reaction against Descartes's view of the self.

But Ryle was not alone in holding such an extreme anti-Cartesian view. He also has influential followers. For example, Anthony Kenny in his *The Metaphysics of Mind*, remarks, "Like Ryle, I regard the inheritance of Descartes as being the single most substantial obstacle to a correct philosophical understanding of the nature of the human mind," (1989: vii). Thus, Kenny rejects both the self as an entity and the first person pronoun "I" as a referring expression (see *Ibid*. 87-88). Similarly, Daniel Dennett remarks, "What is a self? Since Descartes in the seventeenth century, we have had a vision of the self as a sort of immaterial ghost that owns and controls a body the way you own and control your car," (Dennett in Kolak and Martin, eds., 1991: 335; also see Dennett 1991; 1984; 2003). Descartes's notion of the 'soul' is also described as 'a ghost in the machine,' as 'religious or spiritualistic beliefs,' as 'spiritual substance,' etc., (see e.g., Hofstadter and Dennett, eds., 1981; cf. Murphy and Brown, 2007: Ch. 1; Armstrong 1999: Ch. 2; Tye 2003: 133). It has also been said that in the age of cognitive science, the *mind* can only be taken as an aspect of the brain that arises from the complex physical brain mechanism (see e.g., Lakoff and Johnson 1999: 3-10; cf. Varela, Thompson and Rosch, 1991; Bennett and Hacker 2003: Ch. 12).

But the main problem with such reactions against Descartes's view of the self has to do with the fact that they all attack the straw man version of Descartes's position. For example, Descartes never talked about the notion of 'ghost' in any of his own writings. In this regard, Mel Thompson rightly points out that Ryle was solely responsible for portraying the self as a "ghost" and in doing so, for caricaturing what he then takes as the "official" dualist view of the relationship between the mind and matter as "the ghost in the machine". Thompson complains that these days, in philosophy (as in many other spheres of life such as politics), caricatures are more memorable than historical characters. As Thompson further notes, this has been the sad reality of Descartes's view of the self, which is often seen through Ryle's logical behaviorist eyes (2009: 17-18).

Similarly, Descartes never used the term 'soul' for theological/religious reasons contrary to the orthodox view. For example, as Cottingham argues, although Descartes's talk of a non-physical soul is compatible with theology, his belief in the thinking part of him being non-

33

material was not motivated by theological reasons. Cottingham claims that the modern reader may well be misled by the word 'soul' supposing that it is 'spiritual' in the sense of having to do with heightened aesthetic, moral or religious sensibilities. But Cottingham claims that Descartes's 'soul' is by no means limited to the 'soulful' (in the modern sense) aspect of our mind. Rather in Descartes, 'soul' and 'mind' are synonymous and merely function as convenient labels for res *cogitans*—that which thinks. The bottom line here is that when Descartes uses the above terms, he has in mind the whole range of conscious mental activity (1986: 111).

Thompson-Cottingham's remarks above show why widespread anti-Cartesian views of the self are based on serious mischaracterization of what Descartes actually says about his view (Bolton 1994: xx). More importantly, in his own time, Descartes was well aware of the problem of his critics distorting his view of the self. For example, in responding to one of his critics, Bourdin, Descartes remarks:

> Here, as almost everywhere else, my critic represents me…with comments that are wholly at variance with my true views. But it would be too tedious to list all his fabrications….As for what should be termed 'body', or 'soul' or 'mind'…I gave an account of two things, namely that which thinks and that which is extended….However, I did call one of the substances 'mind' and the other 'body'; if my critic does not like these terms, he may employ others, and I shall not complain (AT VII *Seventh Set of Objections with Replies* 487 BB CC: CSM 329).

In bringing this point up, however, I am not implying that Descartes's view of the self is free from problems. As I will briefly explain in the next section, that is not the case. Having said that, the objections raised at the terminological level, do not seem to pose any serious threat to Descartes's view of the self. How about the 'unobservability of the self' objection that was raised at the existence level? Since this is the issue I will discuss in detail in chapter three (in relation to Hume's view of the self) and in chapter seven (in relation to my own view), here my remarks will be very brief.

As I see it, the 'unobservability of the self' objection rules out the self's existence based on a misleading assumption, i.e., for something such as a self to be real, it must be observable. But the question remains: observable in what sense? We commonly use the verb 'to

observe' in a literal sense, that is linking it with *direct observation* on the one hand and in a non-literal sense, i.e., linking it with *indirect observation, on the other hand*. From the standpoint of a literal sense of observation for example, as I write up this thesis, all being equal, I am epistemically justified in making the following inferences:

(i) I observe that the laptop in front of me has a certain size, a certain shape and it occupies a certain region of space, *inter alia*.

(ii) My fingers are typing on the computer keypad as I feel the pressure, the clicking sound, etc.

(iii) What I am typing is appearing on the computer screen, which I can read, delete, etc.

Such inferences in (i)-(iii) are often made under the assumption that we are directly observing objects that fall within our visual field, in this case, the desktop in front of me. There is also what I am calling *a non-literal sense of observation*, according to which the existence of something perhaps can only be inferred on the basis of analyzing its effect. In this case, a non-literal sense of observation is suggested under the assumption that certain things can be observed indirectly. For example, physicists tell us that we do not directly observe atoms, yet we believe in their existence given our knowledge of their properties (e.g., tracking electrons that bounce off atoms). Likewise, we do not have direct access to black holes owing to their immense gravitational pull, from which light itself cannot escape. But the lack of direct access to observe such things is hardly a reason to infer that atoms or black holes do not exist. It is not uncommon in science to postulate the unobservable to account for the behavior of the observable things (see e.g., Green 1999/2003). Of course, here I am not using such analogies from science to defend Descartes's view of the self. Rather, my point is that such analogies from science can help us see the implausibility of assuming (quite in general) that if something is unobservable then it must be the case that its existence must be unreal.

Having distinguished between direct and indirect observation, what can we then say in response to the '*undetectability* or the *unobservability*' of the self objection? As I had said earlier, this objection trades on an ambiguous term, 'observation'. Those who reject

Descartes's view of the self do not make effort at least in what sense they are demanding the observability of a non-physical self, that is, either in a literal or non-literal sense. But since in the literature, Ryle's 'ghost in the machine' is taken as a canonical expression to dismiss Descartes's view of the self, I assume that the non-physical self is denied existence, assuming that it does not pass the observability criterion (taken in the literal sense of observation). If this is what Descartes's opponents have in mind in dismissing Descartes's view of the self, then it remains unclear what force their objection would have against Descartes's position. It seems that Descartes's opponents are relying on Hume's, as we shall see (in chapter three), deeply mistaken assumption that introspection must necessarily entail perceiving the self. For example, having failed to perceive a self via introspection, Hume concluded that there is no such thing called a self (see e.g., a quotation from Hume's *Treatise* under *the existence level* in § 1.1.2.2).

By contrast, Descartes's view of the self can be shown to be compatible with both senses of observation as laid out above. All we need to do here is qualify the sense in which we use observation in relation to self-knowledge. Doing this is extremely important not to find ourselves making a Humean sort of mistake I mentioned above. So how should we go about qualifying observation in relation to self-knowledge? Here I return to my earlier point in § 1.1.1.1. In this section, I argued that Descartes's *cogito* primarily has to do with *de re knowledge*. I also suggested that the best way to capture the sense of *de re knowledge* is by spelling it out in terms of *knowledge by acquaintance*, i.e., unmediated knowledge of the self. So it is open for Descartes without repeating Hume's mistake of requiring 'self-perception, i.e., perceiving oneself as an object,' to say that he has direct knowledge of himself. But by 'direct' here we don't mean that we observe the self as we directly observe objects such as a chair or a table in front of us (cf. Shoemaker 1963: Chs. 1-2). Rather, by 'direct' we mean one's own awareness of oneself immediately without being filtered through anything whatsoever (cf. Moreland 2009: Ch. 5). Even if we do not (perhaps cannot) spell this out in adequate words, the fact seems to remain. That is, self-awareness is an obvious phenomenon. But objections often cited against such claims, by appealing to psychiatric disorders or cases of personality fragmentation, do not stand their ground. I will explain and respond to one such objection in chapter eight. So if the sense of observation is

understood in a way qualified in the preceding paragraph, then Descartes's view of the self can escape the 'unobservability proves the non-existence of the self' objection.

Similarly, there is textual evidence in Descartes's *Second Meditation*, which suggests that Descartes seems to have thought that we can have knowledge of a self via indirect means such as its actions. For example, as Descartes remarks, "But what then am I? A thing that thinks. What is that? A thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions (AT VII 28 CSM: 19). The gist of Descartes's point here is that he knows what a thinking thing is by listing what such a thinking thing does. Put differently, the self's existence can be inferred from its operations. In light of such considerations, the 'unobservability of the self' objection against Descartes is not decisive.

C. *The Unity Level*

It is at the unity level that we find what can be taken as the most serious objection against Descartes's view of the self. Before we see this objection, it would be helpful to recap (D2) from § 1.1.2.2.

> (D2) *X* is an incorporeal substance$=_{df}$ (a) *X* exists; (b) *X* is a non-physical thing: (i) *X* does not occupy space, (ii) *X* does not depend on any material thing, and (iii) *X* is capable of existing without the body, (iv) *X* maintains strict identity over time; (c) *X* is a bearer of mental properties *M*; (d) mental properties *M* depend on *X* for their existence but *X* does not depend on *M* for its existence.

The objection at *the unity level* stems particularly from (D2)-*b's* clauses (i) and (ii). Given (i), it is difficult to see how the self can causally interact with the material body, since it (i.e., the self) does not occupy space. Following the tradition in the literature, I shall call this *the causal interaction problem*. On the other hand, Descartes's statement in (ii) generates what I shall call: *the unity problem*. That is, if the immaterial self does not depend on any material thing, then we wonder what to make of the basis for the unity between them. So the interaction and the unity problems stand or fall together. In the *Second* and the *Sixth Meditations*, Descartes addresses each of these problems, although one can still wonder whether he succeeded in solving them. Here it is not my intention to defend Descartes's

theory of the self against these problems nor am I interested to take any position as to whether or not the above problems are fatal to Descartes's substance dualism.

But most philosophers think that the above problems are insuperable (see e.g., Kim 2005: Ch. 3; Dennett 1991; Atkins 2005: 9; cf. Garrett 1998: 6-9). There are others who do think that the above two problems are not insuperable. So with some modification, they defend Descartes's view of the self (see e.g., Swinburne 1997; Foster 1991; Hart 1988; Robinson 2003). Yet those who deem the *interaction problem* and the *unity* problem as insuperable often insist that self is unreal (see e.g., Garrett 1998: 6-9). On the other hand, those who defend Descartes's view of the self, while acknowledging the above two problems, still embrace Descartes's notion of the *entityhoodness* of the self. For reasons we shall see throughout this thesis, my own sympathy lies with those who retain the notion of the self as an entity.

1.1.2.3 The Lesson and the Challenge

Despite making some positive remarks about Descartes's theory of the self, I do not claim to have shown that Descartes has established the existence of the self. In fact, that was not my aim. Rather, what I have merely attempted to do is demonstrate that Descartes's theory of the self is not as absurd as it is often portrayed in the literature (cf. Hawthorne in Van Inwagen and Zimmerman 2006: Ch. 3). If this is right, then good lessons can be learned from Descartes's theory of the self. One of the lessons I have in mind can be stated in the form of a conditional statement (CS) as follows:

> (CS-1): *If the self exists, then it must be a certain kind of entity, i.e., a substantial entity*.

As we shall see, I will be putting (*CS-1)* within the framework of the Aristotelian metaphysics. It is in light of the Aristotelian metaphysics that I want to defend *(CS-1)* in this thesis. That said, however, describing the self in the third personal language as stated in (*CS-1*) above has its own drawbacks. First, third person language portrays the self as if it is an object, which if it exists, must be found upon searching for it. I already expressed my

dissatisfaction with this sort of Humean requirement. But this requirement still remains to be one of the most popular objections pressed against a defender of the substantial self. I will return to this issue in chapter three. Second, the third person language breeds confusion over what a self is supposed to be.

For example, some philosophers (not to mention psychologists) often talk about the self in terms of locutions such as 'ecological self', 'interpersonal self', 'extended self', 'private self', 'conceptual self' (Keisser in Kolak and Martin, eds., 1991: Ch. 24); 'constructing a self' (cf. Gergen *Ibid.* Ch. 23; McCall in Mischel 1977: Ch. 10); 'immersed self', 'participant self' and 'underself' (see Ganeri 2012: 14); 'the situated self' (Ismael 2007); 'the quantum self' (Zohar 1990); 'multiple selves', 'hierarchical self', 'parallel selves' (Elster ed., 1985: 1-34); 'having a self' and so on. For now, there is no need to analyze all of these locutions. By way of illustrating the kind of confusion I have in mind, I will say something briefly with respect to the last locution, i.e., 'having a self'.

It makes sense when someone says that he/she has a particular personality. For example, we often talk about someone having rude personality versus excellent personality. Taken in either sense, the point is that based on our judgment of the quality of how a person interacts with us or other people, we draw a certain conclusion with respect to his/her personality (see e.g., Goldie 2004; also cf. Maltby, Day, et. al., 2007: Ch. 1). If this is how we are supposed to understand what it means to have a self, then the self seems to have been identified with personality.[20] But those who take the self to be an entity would reject identifying the self with personality as a category mistake. This is because the self is the owner of various sorts of personalities without being itself an amalgamation of those personalities. In this case, the confusion has to do with swapping the self for personality and vice versa. On a similar note, William Alston remarks:

> The homunculus-taboo led many theorists to *identify* the 'self' with a person's cognitions or evaluations of himself, or attitudes to himself. The 'self' is an organized, fluid, but consistent conceptual pattern of perceptions of characteristics

---

[20] I will return to this issue with some detail in chapter eight in relation to some psychological disorders.

and relationships of the 'I' of the 'me', together with values attached with these concepts. It can only breed confusion to identify my self with my conception of myself (Alston in Mischel 1977: 66).

The gist of Alston's remark above is that the misguided [Rylean sort of] fear about homunculi, i.e., a little man inside (i.e., brain) can hardly justify swapping two distinct domains back and forth as if they are identical. But in saying this, neither Alston nor I assume that all forms of everyday talk in which 'self' appears is problematic. For example, Elmer Sprague nicely sums up how we use in everyday talk 'self' as prefix and suffix in such compounds as 'self-conscious', 'self-centered', 'myself', 'themselves', 'I am not myself today' and the like (see Sprague 1978: 25-26). But at the moment, my concern is not with such ubiquitous everyday usage of the term 'self'. Rather what I am interested in is with an ontological implication of the use of 'self' in locutions mentioned earlier (e.g., 'having a self'). In that case, we saw that the third person language opens a door for confusion. So in the interest of avoiding any such confusion as well as for other reasons we shall see, I spell out (*CS-1*) in terms of first personal language as follows:

(CS-2): *If I am a self, then I must be a certain kind of entity, i.e., a substantial entity.*

As we saw, Descartes's official answer to *the substantial-self question* is in line with what is stated in (*CS-2*) above. But there are three other alternative answers to *the substantial-self question* that I want to briefly mention here. The first is Locke's answer. In his *An Essay Concerning Human Understanding* (see Ch. XXVII Nidditch, ed.: 1975), Locke does not commit himself to the view that the self is a substance. Although Locke himself was a friend of substance ontology, as I shall argue, he divorced the notion of the self from substance. In doing so, he strengthened most of the contemporary philosophers' opposition to the ontology of the self. I will pick up this discussion in chapter two.

The second is Hume's answer. In his *Treatise of Human Nature* (see Part I and Part IV 1888 ed., Selby-Bigge), Hume dismisses altogether the traditional notion of substance (associated with Aristotle) in general, and Descartes's doctrine of the self in particular—Hume's own official view being that a self is a bundle of properties. In this case, Hume's view even all the

40

more led some influential philosophers to advocate for a 'no self theory' (e.g., Parfit 1984). I will discuss Hume's view in chapter three.

The third is Kant's answer. In 'Paralogisms of Pure Reason' in his *Critique of Pure Reason*, Kant argues against what he calls rational doctrine/pure psychology, which attempts to establish the subject of judgment (i.e., experience) through non-empirical means such as mere analysis and demonstration (see Smith, ed., A341/B399-A344/B402). So in the 'First Paralogism,'[21] Kant attempts to show why we have no epistemic ground to establish the substantiality of the self (A349-A351). In saying this, Kant is directly challenging the Cartesian view of the self (see also Gardner 1999: 145-148). However, as Patricia Kitcher points out, like Descartes, Kant believes that any cognitive state requires a subject. Yet unlike Descartes, Kant does not conceive of this self as a simple substance, but as a system of syntactically connected or connectable states (1990: 187). In this case, as we shall see in chapter three, Kant's view is anti-Humean.

Of the four answers we have considered so far to *the substantial-self question*, as we shall see, Locke's and Hume's answers are the ones that take center stage in setting the tone for the contemporary philosophical debate on personal identity. By contrast, Descartes's answer is largely overlooked. But so far, I have not said anything about the Aristotelian answer. This is because the notion of the self is not something Aristotle explicitly discussed, although it could be reconstructed from his works (see e.g., Owens 1988: 707-722). Since Aristotle is most well-known, *inter alia*, for introducing the ontological concept of substance, my own interest lies on this aspect of his work. In light of this, in this thesis, one of my main goals is to develop a positive case for substance ontology with an aim to show its relevance both for the ontology of the self/person and the metaphysics of personal identity. In this regard, I will critically examine both Locke's and Hume's views on substance ontology on the one hand, and the contemporary neo-Lockeans' and neo-Humeans' reaction towards substance ontology on the other hand. At this point, it may be asked why

---

[21] According to Kant, "a logical Paralogism is a logical syllogism which is fallacious in form, be its content what it may," (A341/B399).

putting much emphasis on criticizing non-substance ontology approaches is needed. Such an emphasis is needed due to a wide-spread skepticism in the literature that substance ontology can be effectively used to shed light on the problem of the metaphysics of personal identity over time (see e.g., Martin and Kolak, ed., 1993). In this regard, most philosophers' skepticism is rooted in Locke's and Hume's analysis of traditional substance ontology. So in my view, an adequate defense of a positive account of substance ontology requires us first to engage with alternative approaches. I will take up this discussion in chapters two through five. In doing so, ultimately, my aim is to appropriate Aristotle's notion of substance in defending the conditional claim stated in (CS-2) above. That is: *If I am a self, then I must be a certain kind of entity, i.e., substantial entity*. I will situate this conditional claim entirely within the framework of Aristotelian metaphysics. In doing so, my goal is to defend the *entityhoodness* of the self by giving broadly Aristotelian reasons. It is from such discussion that what I shall call: the *Aristotelian answer to the substantial-self question* will eventually emerge. Bringing Aristotle in this way into the contemporary debate on the ontology of the self in general and personal identity in particular has not been sufficiently emphasized.[22] So, I consider my discussion in this regard to be one of the crucial aspects of my contribution to the ongoing debate.

1.1.2.4 Boethius's Way

Up to this point, *the substantial-self question* has been discussed only in relation to the notion of the self. But there is another important notion which I have not yet linked to *the substantial-self question*. That is: the notion of a person. Etymologically, the term 'person' derives from the Latin persona: a mask worn by an actor who plays some kind of role or character. The same meaning applies to its closest Greek cognate προωσπον/*prosopon*. This notion had originated in the Roman law, where persons are perceived to be bearers of legal

_____
[22] For excellent works done in recent years within the framework of broadly Aristotelian metaphysics, see e.g., Lowe (1996), (1998), (2006) and Oderberg (2007).

rights. The notion of the person also became associated with moral value through the influence of Christian tradition (see for details Mauss trans., Halls in Carrithers and Collins, eds., 1985: Ch. 1; Peacocke and Gillett, eds., 1987).

In everyday language, the notion of the person is coextensive with other frequently used terms, namely 'human being' and 'man'. In fact, in ancient philosophy, for example, there is no term for 'person' which is distinct from 'human being' (Gill, ed., 1990: 7). In this case, we can swap the predicates of any of the following three sentences with each other *salva veritate*, i.e., without any loss of meaning:

(A) David Cameron is *a person*
(B) David Cameron is *a man*
(C) David Cameron is *a human being*.

But as we shall see, the liberty we exercise in swapping (in ordinary usage) the predicate terms above in (A)-(C) without risking ambiguity is only temporary. This is because, following John Locke's definition of a person (see below), bioethicists and philosophers argue that the above predicate terms do come apart, i.e., they can be mutually exclusive. In such cases, for example, the notion of the person is often singled out as having unique status over other terms such as 'human being'. Here a good case in point is the current abortion controversies that primarily focus on the question of whether or not a fetus has the status of personhood. But how are we going to settle this issue? The answer directly hinges on what one takes a person to be (see e.g., Oderberg 2000: 174-184; cf. Perry 2002: 199). To make sense of how philosophers usually argue about the notion of the person, first we need to understand how they often tend to think about it. For example, Sprague claims that the term 'person' is not a name that stands for a distinctive kind of thing, as 'frog' or 'diamond' may be. Rather as Sprague claims, we use the term 'person' as a sort of title that we bestow on something if that thing satisfies certain appropriate criteria, as 'doctor' or 'policeman' may be (1978: 61).

Here Sprague speaks for many philosophers. So given Sprague's characterization of the notion of a person, the question remains: Is person a sortal concept? By sortal concepts,

following Wiggins, I mean those concepts, "that present tensedly apply to an individual *X* at every moment throughout *X's* existence," (2001: 30; also see Lowe 1989: 1). For example, the term 'human being' is a sortal concept, since it satisfies Wiggins's principle. By contrast, terms 'president' or 'infant' do not satisfy Wiggins's principle, since these concepts are applied to any given entity temporarily. For example, I was an infant at some point in the past but that concept no longer applies to me. Currently Barack Obama is the president of the United States. But after his second term in the Oval Office, the title 'president' will no longer actively apply to him. Wiggins calls the latter sorts of concepts *phased sortals* whereas he calls the former (i.e., human being) *substance sortals*. Wiggins claims that only substance concepts give a fundamental kind of answer to the question 'what is X?' (*Ibid*.). In light of this, Sprague's characterization of the notion of the person above forces us to consider 'person' only as a phase sortal. Others like Christian Kanzian (in Gasser and Stefan 2012: Ch. 11) argue that the term 'person' is an incomplete sortal, which must depend on other full-fledged sortals like a 'human being' for it to be a complete sortal. Such view of personhood is what I very much want to dispute in this thesis.

However, to get a clear picture of why most philosophers take the notion of a person as phase sortal, it would be helpful to look at some questions that P.F. Snowdon (in Lovibond and Williams, eds., 1996: 36) suggests any account of a person must address:

(1) Is the term 'person' ambiguous, or can the uses of it be brought under a single concept?

(2) Is the term 'person' one in which if an object satisfies it at any time during that object's existence then the object must satisfy it at all times? If not, can we make sense of the idea of a pre-existent object of some type becoming a person, or something which is, at some time, a person ceasing to be a person?

(3) In our elucidation of the term 'person' we often refer to psychological features, but how should we pick out those features?

As to (1), there is no doubt that the term 'person' is ambiguous, since it is used to designate human persons, artifacts (i.e., robots, computers, corporations), extraterrestrial persons (if there are any), divine persons such as God, angels, demons and non-human animals such as apes and dolphins (see e.g., Puccetti 1968; DeGrazia in Singer, ed., 2006: Ch. 3). So those

who take the notion of the person to fill different roles or functions argue that there cannot be a single notion of a person (see e.g., Rorty and Morton in Gill, ed., 1990: Chs. 1 and 2 respectively). So in the context of the present discussion, we disambiguate the term 'person' by adding an adjective in front of it and describe it as 'human person'. Unless otherwise indicated, in this thesis, by 'person' I always mean 'human person'.

That said, how should we go about answering the questions in (2) above? In the course of our discussion, I will suggest my own answers to the questions. But my main discussion in this regard will come in chapter eight, where I will examine the answers given to (2) by three influential philosophers, namely Singer (1993), Tooley (1972) and Dennett (in Amelie Rorty 1976). These philosophers (including many others who follow them), typically answer the questions in (2) above as follows: personhood is not something an entity has, for instance, since the time of conception. Rather, personhood is something an entity gradually comes to have after a period of some developmental stage (e.g., infancy). These philosophers also claim that personhood is something an entity could lose altogether under certain circumstances that have to do with amnesia, senility, being in a comatose state, etc.

At this point, the question that comes to mind is this: what properties are essential for an entity such that, were an entity to lose them then that entity becomes no longer a person? On the flip side, a question can also be asked: what properties must an entity have in order for it to be able to continue as a person? These are questions that the majority of philosophers attempt to settle by answering question (3) above (see e.g., Noonan 2003). They do this mainly by appealing to Locke's definition of a person, according to which, "a person is a thinking intelligent Being, that has reason and reflection; one that can consider it self as it self; and the same thinking thing in different times and places; which it does only by that consciousness, which is inseparable from thinking," (*Essay*, XXVII, 9: 335).

As we shall see, my own view as to what constitutes personhood stands sharply opposed to that briefly highlighted above. I will raise serious objections against Singer's, Toole's and Dennett's conception of personhood. I will do this by situating my objections in the current discussion of the metaphysics of dispositions (see e.g., Bird 2007; Martin 2008; Heil 1998,

2012). As we shall see, Singer, Tooley and Dennett seriously conflate, among other things, the manifestation of dispositions with the entity that manifests those dispositions. In other words, they fail to see the distinction between an object that manifests certain dispositions on the one hand and the manifesting of dispositions on the other. They also mistakenly think that an object's failure to manifest certain dispositions is evidence for the object's non-existence. These are profound confusions and once we get a good grip on the nature of the confusions related to personhood, we can see clearly why the idea that personhood begins in conception (i.e., fertilization) is defensible. One of the contemporary philosophers who rigorously defends the view that personhood begins in conception is Oderberg (see e.g., Oderberg 2000).

I have said enough by way of showing how the concept of a person is understood in the context of the broader literature.[23] But as I indicated at the beginning of this section, my goal is to link the concept of a person to *the substantial-self question*. To do this, I want to understand the notion of a person as an entity representing concept in the same way that I did with the notion of the self (see e.g., § 1.1.2.3). That is, in the words of Chisholm, "persons are *entia per se* [i.e., real things] and not *entia in alio*—not ontological parasites like shadows," (1989: 57). But what sort of conception of person would capture Chisholmean assumption? Here I suggest the medieval conception of personhood as described by Boethuis (480-526 AD). In one of his theological tractates, Boethuis defines persona as follows: "*a person is an individual substance of a rational nature*" (H. F. Stewart, trans. 1918: 85). Here Boethuis' definition of a person is rooted in classical notions of substance and rationality (see Gill, ed., 1990: 7). In this case, Aristotelian substance ontology takes center stage. For example, Thomas Aquinas (1225-1274 AD) defended Boethuis's conception of person within the framework of Aristotelian substance ontology (see e.g., I Sentences, d. 25, q. I, Mandonnett, ed., Vol. I, 600-606; also see Wippel 2000: 24). More importantly, Boethuis himself had written a commentary on Aristotle's *Categories* (see *In Categorias Aristotelis libir quattuor*. Migne PL 64). In light of such

_____

[23] For a detailed account of the definition of a person from historical, theological as well as contemporary philosophical standpoints, see Jenny Teichman (1985: 175-185).

considerations, it is not difficult to see Aristotle's influence on Boethius' view of a person.[24]

So following Boethius and Aquinas, I too want to defend the notion of a person as a *substantial entity*.[25] That said, as I did with the notion of the self in § 1.1.2.3, here too I spell out the notion of the person in a conditional statement using first person language as follows:

(CS-3): *If am a person, then I am a certain kind of entity, i.e., a substantial entity.*

In light of the hitherto considerations, we can rephrase *salva veritate the substantial-self question* as *the substantial-person question*, i.e., Is the person a substance? That means that we can use either formulation of the question to get across the same point, that is, the substantiality of the self/person. So in this thesis, I will use the terms 'self' and 'person' interchangeably.[26]

What I have tried to do up to this point has been to underscore the importance of taking the self/person as an entity. From now on, I will locate my discussion of *the substantial-self question* or *the substantial-person question* within the framework of the contemporary debate on personal identity. I now turn to that discussion.

1.2 Historical Context

In their *Naturalization of the Soul: Self and Personal Identity in the Eighteenth Century*, Martin and Barresi point out two significant revolutions that took place in personal identity theory. As Martin and Barresi describe:

> In the eighteenth century in Britain, there was a revolution in personal identity theory. In our own times, beginning in the early 1970s, there has been another. It is well known that in the earlier revolution, the self as immaterial soul was replaced

---

[24] Boethuis' view of personhood is mainly motivated in getting clearer about the Christian Doctrine of Trinity and the person of Christ (see e.g., Marenbon, ed., 2009: 107, 112-13, 118-24, 145-51, 158, 164-5, 175).

[25] Aquinas's view of personhood raises some complications that I am not interested to go into. For example, Aquinas believes in composite person (i.e., body and soul). But he also seems to hold that soul cannot be complete without the body. Scholars fiercely debate on the details of Aquinas's view see e.g., Toner (2009: 121-138). But nothing I will say here hangs on such controversies.

[26] Locke also uses these terms synonymously, see *Essay* XXVII; also see Shoemaker (1963: Ch. 1); Lowe (1996; 2008). Others oppose such synonymous usage see e.g., Abelson (1977: Ch. 6).

with the self as mind. This replacement involved movement away from substance accounts of personal identity, according to which the self is a simple persisting thing, toward relational accounts of personal identity, according to which the self consists essentially of physical and/or psychological relations among different temporal stages of an organism or person (2000: ix).

Again with respect to the contemporary revolution, Martin and Barresi say:

> At the heart of the revolution in our own times has been the emergence of two questions where previously, it seemed, there had been only one. The traditional question is: What are the necessary and sufficient conditions for personal identity over time? That is, what must obtain in order for the same person to persist, and what, if it does obtain, guarantees that the same person persists? The new question is: What matters fundamentally in a person's apparently self-interested concern to survive? That is, from the perspective of what, in normal circumstances, would count as a person's self-interested concern to survive, is it fundamentally personal identity that matters—does the person fundamentally want to persist—or is what matters other ways of continuing that do not themselves suffice for identity? (*ibid*.).

The impact of each of these revolutions in shaping up the direction of the contemporary debate on personal identity is undeniable.[27] But the question that still remains is, whether these revolutions have contributed towards our understanding of the issue of personal identity in a positive way. To see this, it is crucial to make some observations with respect to the changes that came about as a result of these revolutions (hereafter the 18th century revolution and the 1970s revolution, respectively). First, as we can see from the first quote above, the 18th century revolution was responsible for: (i) the beginning of the

---

[27] See the details on the development (and outcome) of these revolutions from a historical aspect in Martin and Barresi (2000). For a more comprehensive account of the history of the development of personal identity and the self that goes all the way back to the ancient Greece, see again Martin and Barresi, *The Rise and Fall of Soul and Self* (2006). Whilst these two works give an impressive and rigorous account of the history of personal identity and the self that spans well over 2000 years, as will be made clear in this thesis, I remain very much opposed to the conclusions these authors seem to draw with respect to the nature of the self and of personal identity as a result of their historical studies. For example, both of these authors reject the substantial view of the self, as no longer a viable (or at the least deeply problematic) view to advocate. It is worth comparing Martin's and Barresi's position to that of Richard Sorabji's, who also has done an extensive study on the nature of the self and personal identity. In his *Self: Ancient and Modern Insights about Individuality, Life and Death* (2006), Sorabji defends a positive view of the self within the framework of the notion of embodiment. He also thinks that the notion of the self is capturable via 'I' thoughts. He rejects the Humean/Parfitean account of the self, which reduces the self to the sum of a bundle of properties. Also see Jerrold Seigel's *The Idea of the Self* (2005), sophisticated and extremely detailed account of the self since 1600 to present. Seigel gives *inter alia* thorough critique and response to post-modern notion of the autonomous-egocentric and disengaged view of the self. Also see, Udo Thiel's discussion of personal identity from *Descartes to Hume* (2006).

abandonment of 'substance ontology' in personal identity theory; and (ii) the beginning of the establishment of the 'relational account of personal identity'. As we shall see in chapter three, in the case of both (i) and (ii) above, John Locke's influence is immense. Second, the 1970s revolution first and foremost should be understood as being responsible for the considerable attempt made by philosophers to establish (ii) above, as the most preferred view of personal identity against that of the substance account of personal identity.

More importantly, the 1970s revolution was also responsible for the subsequent attention given to what Martin and Barresi describe as (in the second quote above), the traditional question, which is concerned with providing a necessary and sufficient condition for personal identity over time on the one hand and the notion of whether or not numerical identity is what matters in one's own survival on the other. In the latter case, Parfit's influence excels any other contemporary personal identity theorist (see e.g., Martin 1998).

Although I agree with much of Martin's and Barresi's observations regarding the effects of the two revolutions briefly discussed above, I remain unpersuaded with their emphasis on singling out only two questions (i.e., the traditional question and the what matters in survival question) as defining the nature of the contemporary debate on personal identity. In this chapter, I shall try to show why I think Martin's and Barresi's observation is short-sighted. Contrary to many philosophers, I hold that the most central question in contemporary personal identity theory is neither the 'traditional question' nor the 'what matters in survival question,' but what I call the 'ontological question' or what I called earlier the *Aristotelian question*. In light of what I will have to say in defense of the *Aristotelian question*, this thesis will insist that the announcement often made by many contemporary philosophers with respect to the irrelevance of substance ontology for personal identity is not only premature, but ultimately, it is an implausible position to adopt. I will push one step forward and defend in this thesis the claim that currently the most preferred view, i.e., 'relational/psychological account of personal identity' is unsuccessful in advancing our knowledge of our identity as persons. I will argue that what is often considered as progress in the 'relational account of personal identity,' is something

49

that is based on questionable methodology,[28] which mainly appeals to various sorts of puzzle cases. So if what I will say in this regard turns out to be right, then the most preferable and defensible view of personal identity and the self/person will indeed be the one that is spelled out within the framework of substance ontology. In light of this, the main goal of this thesis is to defend the importance of substance ontology for personal identity. Throughout this thesis, therefore, I will be giving several interconnected but distinct arguments that flag out the centrality of substance ontology for our understanding of personal identity on the one hand and the nature of the self/person on the other. This means that I will be rejecting the 18th century revolution's move away from substance account of personal identity to relational account of personal identity. I hope to show in this thesis why my rejection in this regard is justified.

1.2.1. The Common Answer

That said, however, as Martin and Barresi have observed, the contemporary philosophical debate on personal identity very much focuses on *the traditional question*. Recall that the traditional question concerns with providing a necessary and sufficient condition for personal identity over time. So if we were to ask the question: What precisely is at the heart of the contemporary philosophical debate on personal identity? The most common answer we would get is a 'criterion of identity' for persons.[29] This notion was introduced in philosophical discussion by Gottlob Frege.[30] In the *Grundlagen*, Frege remarks, "If we are to use symbol *a* to signify an object, we must have a criterion for deciding in all cases whether *b* is the same as *a*, even if it is not always in our power to apply this criterion," (1884: § 62 trans by Austin 1953; see also Dummett 1973/1981). Frege gave an example for such

---

[28] My own account of how to best utilize puzzle cases/thought experiments, in this case, will be discussed in chapter four.

[29] See e.g., Perry (ed.) (1975); Rorty (ed.) (1976); Noonan, (1989) and (2003); (1993) (ed.), 2009; Williamson (1986); Parfit (1984: Part III); Swinburne and Shoemaker (1984); Kolak and Raymond, ed. (1991); Baillie (1993a), (1993b); Oderberg (1993); Sider (2000b); Olson (2010); Korfmacher (2010) and Gasser and Stefan (ed.) (2012). What I will say here regarding this notion is very sketchy. A detailed treatment of this notion awaits Chapter five.

[30] However, Lowe claims that Frege deserves credit for having introduced the terminology so much more than the notion of the criterion of identity itself, since the notion of the criterion of identity was already understood by Locke long before Frege (see, e.g., Lowe 1989b: 2). It was also Locke who first coined the term, 'sortal' (see *Essay* III.III.15: 417).

criterion. As he sates, "The judgment 'line *a* is parallel to line *b*'… can be taken as an identity. If we do this, we obtain the concept of a direction, and say: 'the direction of line *a* is identical to the direction of line *b*'''" (Frege 1884: § 64). Frege's point is that parallelism is a criterion for identity between two lines. The notion of criterion of identity is also at the heart of the contemporary discussions besetting the problem of personal identity. In this regard, John Locke was the first and arguably the most influential historical figure who set the agenda for the problem of personal identity in its modern form.[31] For example, in his *Essay Concerning Human Understanding* Book II, chap. XXVII (27), Locke posited this question: "wherein personal identity consists," (*Essay*, chap. 27: 9).

In asking such a question, Locke was after, what the contemporary philosophers describe as, the diachronic identity of persons. This is the notion that pertains to identity over time. This aspect of identity is often contrasted with synchronic identity, which concerns identity at a time.[32] Of these two aspects of identity, diachronic identity is the one that dominates the contemporary discussion on personal identity.[33] As we shall see, it is also the one that proves to be difficult to get a handle on in relation to the problem of diachronic personal identity. But is there a problem of personal identity? Many will answer this question with a resounding 'yes'. But if we follow David Lewis's lead here, our answer must be 'no, there isn't any problem'. As Lewis remarks:

> Identity is utterly simple and unproblematic. Everything is identical to itself; nothing is ever identical to anything else except itself. There is never any problem about what makes something identical to itself; nothing can ever fail to be. And there is never any problem about what makes two things identical; two things never can be (Lewis 1986: 192-193).[34]

I take the gist of Lewis's claim here to be the trivial logical truth that can be expressed formally as: $(\forall x) (x=x)$. That is, for all *X*, *X* is identical to *X*. In other words, identity is

---

[31] We will look at Locke's theory of personal identity in chapter two.

[32] I will make some relevant distinctions in regards to the notion of identity in § 1.2.2.

[33] Synchronic identity itself is a subject of considerable controversy. For example, since the unity of consciousness presupposes unity and multiplicity at a given time, it has been said that cases such as *multiple personality disorder* and *split brain* pose a challenge to synchronic identity. This objection directly comes face to face with the claim I will be defending in this thesis, namely the unity of the self both in the synchronic sense as well as in the diachronic sense. I will respond to this particular objection in chapter eight.

[34] See also Kripke (1972/1980: 3-5); Salmon (2005:153-154); Lowe (2002: Ch. 2).

reflexive, inter alia. This reflexive identity relation also obeys Leibniz's law of *indiscernibility of identicals*, according to which if *X* is identical to *Y* then whatever is true of *X* must also be true of *Y* and vice versa. I will return to this issue in some detail in chapter five. In chapter six, I will rely on Leibniz's law of *indiscernibility of identicals* to defend the claim that the person/self is distinct from the physical body that embodies it. Leaving that aside for now, it is clear that as Lewis sees it, self-identity holds of necessity, i.e., it could not be otherwise. Formally the notion of self-identity can be stated as follows: $(\forall x)\ \Box(x=x)$. That is, for all *X*, necessarily *X* is identical to *X*.

For some philosophers, however, this feature of identity may come across as mystifying, since it entails the relation everything has to itself and nothing else. To put it formally: $(\forall x)$ (Rxx). In this regard, the question we face is this: 'how can a thing be said to stand in relation to itself'? Such a relation seems to be difficult to make sense of. In common parlance, however, we make perfect sense of the difference between the following two statements:

      A. John got mad at Smith
      B. I am mad at myself

What (A) says is that, 'John stands in an angry sort of way in relation to Smith'. Similarly, given that the person who is uttering the statement is making first personal assertion, it seems unproblematic to understand (B) also to mean, 'I stand in an angry sort of way in relation to myself'. Moreover, the nature of the relation in (A) is *extrinsic*, i.e., one object standing in relation to another object whereas in (B) the relation is *intrinsic*, i.e., no outside object other than oneself, is standing in relation to oneself. In the latter sense, therefore, the relation is internal to oneself (cf. Cameron in Poidevin, Simons et al., eds., 2009: 267). Given that the relation in (B) is internal, one may object to it in the following way. That is, in the case of (A) it seems easy to determine the *relata* in the predicate 'got mad at' (e.g., John and Smith). But in case of (B), it is hard to establish the *relatum* without running into circularity. For example, as Noonan remarks, "numerical identity can be characterized, as just done, as the relation everything has to itself and to nothing else. But this is circular, since "nothing else" just means "no numerically non-identical thing" (2009). Here two

questions come to mind: one has to do with whether such circularity is vicious and the other has to do with whether such circularity can be avoided. For reasons we shall see, I take the circularity here to be both benign and unavoidable (cf. Hawthorne in Loux and Zimmerman 2003: 99).[35]

Following Lewis's lead, Noonan recently argued that the traditional problem of personal identity should be reformulated as the question that is not about identity (see Noonan in Gasser and Stefan 2012: Ch. 4).[36] As it stands, it seems to me that Lewis's insistence on the unproblematic nature of everything being necessarily identical to itself seems to be entirely accurate. The onus, therefore, must be on anyone who thinks otherwise, to show us how and why anything can fail to be identical with itself.[37] Assuming that Lewis is right in saying that there is no problem about identity, in what sense then diachronic identity may be taken to pose a problem, if not in the sense of the very notion of identity? One way we can answer this question is by looking at the issue from the standpoint of, *inter alia*, classification and reidentification, i.e, identification over time. Diachronic identity for objects in general and human persons in particular, presupposes that objects endure through time despite undergoing various changes.[38] But the question is how do we go about picking out objects over time? Here is where we begin to see the role and place of the notion of the criterion of identity in personal identity discussions. For example, T.E. Wilkerson remarks that when we claim to have identified an *X*, we mean either we have discovered that it was an *X* as opposed to a *Y* or a *Z*, or we may be claiming that it was the same *X* that we picked out in the past (Wilkerson 1974: 22; Cf. also Rorty (ed.), 1976: introduction). In this case, as Wilkerson further points out, we face two kinds of questions:

---

[35] That said, as far as the reflexive identity relation goes, it may appear to be, as Lowe remarks, an odd feature of relation, which can only obtain between a thing and itself. Perhaps the best we can say here is that we have an unusual relation. But that does not give us any plausible reason to deny that we have a genuine relation (2002: 23).

[36] In light of Lewis's dictum, Noonan's article focuses on indeterminacy in personal identity as opposed to the source of indeterminacy being in identity itself.

[37] One may object to such claim inter alia by appealing to quantum mechanical effects regarding the identity of electrons in the superposed state/quantum entanglement. But not everyone agrees that quantum mechanical effects pose a threat for self-identity applied to persons or substances (see e.g., Heil 2012: 44-52; Stairs in Kolak and Martin 1993: Ch. 28).

[38] This is a central point, to which I will return in due course.

(a) What kind of thing is an *X*?; and (b) How do we decide whether this *X* is the same as that? (*ibid*.). The former question deals with classification. Here an example would be: 'the object I saw under the freeway bridge last night was a homeless man, not a dog or a rock.' The later question deals with reidentification. An example of this would be: 'this man is identical with the homeless man I saw last night under the freeway bridge.' Wilkerson characterizes types of the first and the second example, in terms of 'statements which classify' and 'statements which reidentify', respectively (*Ibid*.).

Wilkerson also rightly emphasizes that the distinction between classification and reidentification needs to be qualified. Firstly, classification presupposes reidentification. That is, once we classify things, we need to think of those things as persisting in order to reidentify them at a later time (*Ibid*. 23).[39] Secondly, reidentification presupposes classification. For example, we can't classify something for the second time unless we have already classified it for a first. Thirdly and more importantly, in classifying the kind of things which may be reidentified, one is necessarily counting those things on the basis of using concepts which have criteria of distinctness built into them (*Ibid*. 23). As Wilkerson nicely sums up:

> For example, in classifying a thing as a cabbage I am not merely distinguishing one part of the world from the rest of the world, I am also counting the number of cabbages in this bit of the world. In contrast, in classifying a thing as red I am not counting anything, for the concept of red does not on its own yield any principle of counting…[thus] one should not suppose that the problem of counting arises only when one is asking questions about reidentification. Clearly if the question of reidentifying Xs is to arise at all, then the concept of X must yield a principle of counting (as 'cabbage' does and 'red' does not); but if the question of reidentification has arisen, then the problem of counting has already arisen with the question of classification (*Ibid*. 23).[40]

The upshot of the above brief discussion on classification and reidentification in relation to personal identity requires us to look for the principles which govern how we count persons and trace their careers over time (Lowe 2000: 272). It is at this point that philosophers claim

---

[39] Wilkerson draws this point from Immanuel Kant's *Critique of Pure Reason*.
[40] For more discussions, see Lowe (1989a); Strawson (1959: Ch. 1); Searle (1969: 85-94).

that we need a criterion of personal identity.[41] It is also a good place for us to see some of the characterization of the notion of 'criterion of identity' before we see its application to diachronic personal identity. But before I introduce some of these characterizations, it is key for us to understand first how in the context of the present discussion, philosophers use the term 'criterion'. In this case, the term 'criterion,' could be viewed from two standpoints:

First, 'criterion' can be used to answer the question: "what evidence of observation and experience can we have that a person P_2 at t_2 is the same person as a person P_1 at t_1?" (Swinburne, in Shoemaker and Swinburne 1984: 3). Let us call this: *the evidential criterion question*. Here an example would be (among other things) the first person memory, where one distinctly remembers having done something in the past. In this case, one can use the first person memory as evidence for his/her persistence over time. Second, the term 'criterion' is also used as a metaphysical principle to establish personal identity over time. The second sense of *criterion* is often described as constitutive, the metaphysical-cum-semantic criterion of personal identity (Noonan 2003: 2).[42] When philosophers debate about the problem of personal identity, what they have in mind is the second sense of 'criterion'. So it is also in the second sense of 'criterion' that we need to understand the following characterizations of criterion of identity.

Schematically, Lowe characterizes the 'criterion of identity' as follows: "if X and Y are Φ, then X is identical with Y if and only if X and Y satisfy condition CΦ" (Lowe 1989a: 23). Lowe's slightly modified version of the same schema runs: "if *X* and *Y* are things of kind *K*, then *X* is identical with *Y* if and only if *X* and *Y* stand in the relation R_k to one another," (Lowe 2000: 272). It seems that Lowe's schema can be applied to establish the identity of objects in general. In this sense, Lowe's schema allows for wider application. Olson's schema runs: "necessarily, if X is a human person at time *t* and Y exists at another time *t**, *X=Y* if and only if…," (Olson in Gasser and Stefan 2012: 47). Similarly, following Olson, Hud

---

[41] However, whether we can come up with an informative criterion for personal identity is a subject of intense controversy, to which we will return in due time.

[42] In saying this, however, we are not assuming that the evidential criterion is irrelevant. As we shall see, it plays an important role in its own right.

Hudson remarks, "one perplexing philosophical puzzle concerns how best to complete the following sentence: Necessarily, if *X* is a human person at a time, *t*, and *Y* is a human person at a distinct time, *t**, then *X=Y* if and only if and because_____,"(Hudson in Gasser and Stefan 2012: 236). Hudson also reformulates criterion of diachronic personal identity that is amenable for 'perdurantists' [43] theory of persistence over time: Necessarily, if *X* is a momentary stage of a human person at a time, *t*, and *y* is a momentary stage of a human person at a distinct time *t**, then *X* is a momentary stage of one and the same human person as *Y* if and only if and because_____, (Hudson in Gasser and Stefan 2012: 236).

In light of the Lowe-Olson-Hudson schemata, the contemporary major controversies surrounding personal identity boils down to specifying conditions for personal identity over time. I shall call this the problem of establishing conditions of diachronic personal identity. By specifying such conditions, most (if not all) personal identity theorists,[44] hold that the blank at the end of the biconditional (in the above schemata) can be filled out with an appropriate non-trivial answer. But what constitutes an 'appropriate non-trivial answer' in this context? As we shall see, reaching a consensus on this point proves to be exceedingly difficult, to say the least. Yet on this very issue philosophical debates have continued unabated. From time to time we see significant efforts being made by philosophers to establish a criterion of identity for persons. The question is: have these efforts been successful? As we shall see, the answer for this question is not encouraging.

I hope that the hitherto brief discussion has given us some highlights of the issues that take center stage in the contemporary controversy on personal identity. In a nutshell, the issue comes down to answering one key question. That is: "what are the logically necessary and

_____

[43] Perdurantists hold that a concrete particular is made up of distinct temporal parts that exist at various times. By contrast, endurantists hold that a numerically one and the same concrete particular exists at different times. For details, see, Loux (1998: Ch. 6); Hawley (2001: Chs. 1 & 2); Haslanger and Kurtz, eds., (2006); and Kanzian ed., (2008). I will have something to say about these issues in relation to substance ontology.

[44] Here I have in mind, in particular, those who identify themselves as defenders of the relational account of personal identity.

sufficient conditions for a person P_2 at time t_2 being the same person as a person P_1 at an earlier time t_1?" (Swinburne in Shoemaker and Swinburne 1984: 3). As we recall, this was what Martin and Barressi earlier called *the traditional question* (hereafter *the persistence question*).

1.2.2 Common Approaches

Views of diachronic personal identity that address *the persistence question* are customarily divided into two main categories, namely *the complex views* and *the simple view*.[45] According to the complex views, personal identity over time consists in some further facts, i.e., other than itself. By contrast, according to the simple view, personal identity does not consist in any further fact other than itself (Gasser and Stefan 2012: 3). So the dispute between the complex views and the simple view with respect to diachronic personal identity mainly hinges on what the locutions, i.e., 'some further fact' and 'no further fact,' amount to. To make sense of these locutions, first we need to link them to the notion of the criterion of identity. As we recall from our brief discussion in § 1.2.1, the criterion of personal identity specifies logically necessary and sufficient conditions for the persistence of persons/selves over time. That is, necessarily, if X is a human person at time t_1 and Y exists at another time t_2, $X=Y$ if and only if_____.

So how should one go about completing the blank at the end of the above *bi*conditional? This is precisely the main point at which theorists of the complex views and theorists of the simple view disagree with each other. But before I briefly explain this disagreement, I want to make some distinctions between various aspects of the notion of identity, which are often confused in discussions related to personal identity. I also want to explain the relation that exists between notions such as 'persistence', 'change' and 'identity'.

A. Identity: Numerical and Qualitative

---

[45] Parfit was responsible for introducing this distinction (see in A. Montefiore, ed., 1973). Recently, Olson argued that no such sharp boundary can be drawn between these views see Olson (in Gasser and Stefan 2012: Ch. 2). Olson may well be right in this regard, but for present purposes, I will stick to the existing tradition in maintaining the distinction.

In its ordinary every day usage, identity usually pertains to how one conceives of oneself. For example, I may view myself in terms of different categories such as Ethiopian-American, male, graduate student, father, husband, and suchlike (see e.g., Ludwig 1997). For present purposes, I set aside such considerations of identity, since for now, my interest is to look at the concept of identity that derives from logic.[46] In this regard, logicians tell us that the concept of identity is an *equivalence relation*, namely reflexive, symmetrical and transitive relation. As we recall, in § 1.2.1, I briefly discussed reflexive relation: $(\forall x)$ $(x=x)$. The remaining two relations are symmetrical relation (SR) and transitive relation (TR). Formally these relations can be described as follows:

(SR): $(\forall x)$ $(\forall y)$ $[(x=y) \rightarrow (y=x)]$, (i.e., for all *X* and for all *Y*, if *X* is identical to *Y* then *Y* is identical to *X*).

(TR): $(\forall x)$ $(\forall y)$ $(\forall z)$ $[(x=y \,\&\, y=z) \rightarrow (x=z)]$, (i.e., for all *X*, for all *Y*, for all *Z*, if *X* is identical to *Y* and *Y* is identical to *Z*, then *X* is identical to *Z*).

Here the identity statements themselves are distinguished from all others by one important property, i.e., they always obey what is commonly known as the Leibniz' Law or the law of *the Indiscernibility of Identicals* (§ 1.2.1).[47] Leibniz's Law (LL) can be symbolized as follows:

(LL): $(\forall x)$ $(\forall y)$ $[(x = y) \rightarrow (P) (Px \leftrightarrow Py)]$, (i.e., for all *X*, for all *Y*, if *X* is identical to *Y* then for any property *P*, *X* has *P* if and only if *Y* has *P*).

---

[46] In saying this, however, I am not assuming that an ordinary way of conceiving our identity has nothing to do with philosophical concept of identity. For discussion of philosophical notion of identity in relation to non-philosophical notions of identities such as social identity (see e.g., Williams in Harris 1995: Ch. 1). Since my goal in this thesis is to discuss the notion of identity from the standpoint of personal identity, I do not intend to get into the contemporary discussion on identity *simpliciter*. In this regard, see e.g., Munitz, ed., (1971); Griffin (1977); Geach (1980); Noonan (1980); Noonan, ed., (1993); Brody (1980); Oderberg (1993a); Hirsch (1982); Williams (1989); Gallois (1998); Wiggins (1980/2001; 1967); Hawthorne in Loux and Zimmerman (2003: Ch. 4); and Parsons (2000).

[47] This law should not be confused with what is generally regarded to be a very controversial law known as: *identity of indiscernibles*, according to which if any given two things share identical (i.e, similar) properties, then those things are identical (i.e., one and the same thing). Formally this principles reads: $(\forall x)$ $(\forall y)$ $[(P) (Px \leftrightarrow Py) \rightarrow (x=y)]$. See Max Black's famous universe with two qualitatively indiscernible spheres, which he used it as a counterexample against *identity of indiscernibles* (1952). For a defense of the identity of *indiscernibles*, see Brody (1980: 7-20). For summary of contemporary controversies on this issue, see Foster (2010). On the other hand, sometimes it has been said that Leibniz's first law itself would be false in *intensional context* in which coreferring terms could not be substituted *salva veritate*. But this objection is entirely out of step with Leibniz's first law. For very good discussion on this see Jubien (1997: Ch. 4); see also Wiggins (2001: Chs. 1-4).

According to Leibniz's Law, for example, if Mihretu and John are identical, then Mihretu and John share one and the same properties. In this case, what we have is not two different entities—Mihretu and John. Instead what we have is one and the same entity, i.e., Mihretu is John and vice versa. Here the 'is' in the statement 'Mihretu is John' implies strict numerical identity. But numerical identity itself must be distinguished from another relation that goes by the same term 'identity' or what is sometimes known as '*qualitative identity*' or '*exact similarity*'. For example, if we have two red and round discs which share all and only the same properties, they would still be two discs and not one. This is precisely because an individual thing like a disc is not exhausted by its properties (Moreland 2001: 21). Here we can understand the relation that holds between two qualitatively similar (but numerically distinct) discs as tokens of the same type.

But when it comes to personal identity, first and foremost we are interested in numerical identity as opposed to mere qualitative similarity. In other words, we want to know whether or not qualitative change is compatible with numerical identity. For example, what happens if a banana changes its colour from green to yellow over time? Do we still have the same (i.e., numerically identical) banana? If the answer to this question is 'yes,' then such an answer would imply that one and the same banana survived changing its colour from green to yellow. This means that numerical identity is entirely compatible with qualitative change but the converse is not true (cf. Shoemaker in Shoemaker and Swinburne 1984: 72). Philosophers also distinguish between synchronic and diachronic identity. That is, X and Y are said to be synchronically identical, just in case X and Y are numerically identical at any given time; whereas X and Y are said to be diachronically identical just in case the numerical identity holds between them over time. But in making such distinctions, we are not assuming that there are different kinds of identity. As Olson rightly reminds us, identity does not come into two kinds, namely synchronic and diachronic. Instead all we are saying here is that there are two kinds of situations where we can ask how many people there are. That is, the synchronic situations involving just one moment whereas diachronic involving a stretch of time (Olson 2010). Along similar lines, Lowe also claims that identity is univocal, that is, there are not different kinds of identity for different kinds of things, any more than there are different kinds of existence. Instead, what we have is different criterion of identity

(Lowe 1989a: 22). Having made such distinctions, I now briefly explain in the next section, how the three notions, namely 'persistence', 'change' and 'identity' are interrelated.

B. Persistence, Change and Identity

Suppose that you had met me six years ago. Back then you would have noticed that I had full hair on my head. You would also have noticed that after six years, I have lost half of my hair. The question then is: am I bald? Not necessarily. This is because, the term 'bald' in the statement such as 'Mihretu is bald,' is a vague predicate. In my case, for example, I can ask: how much hair must I lose to be 'a bald person'? In answering such a question, we inevitably come to face a borderline case (see Keefe 2000; Sorensen 1997/2012; Williamson 1994). But for now, I ignore matters related to vagueness or indeterminacy. Instead my point here is that in losing my hair, I have undergone a qualitative change. In contemporary philosophical parlance: *I had the property of full hair at t_1* but *at t_2* (having lost half of my hair)*, I have gained a new property, i.e., *the property of half hair* (cf. Plantinga 1974). In Aristotle's language, this is an instance of numerically one and the same object receiving 'contraries,'[48] i.e., incompatible properties over time (see *Cat*. 4a10: Ch. 5). Once again, as pointed out earlier in §1.2.2 A, numerical identity is compatible with qualitative change (cf. Smith and Oaklander 1995: Ch. 5).

So at least initially, most of us grant that objects persist through some qualitative change.[49]

––––––––––––––––––––––
[48] But such contraries cannot be had by an object *at a time* for that would entail contradiction. There are some controversies among philosophers on this issue, for example, see the next footnote.

[49] However, we should keep in mind that the kind of reasoning sketched out here does not necessarily apply in the case of other sorts of changes, each of which can be traced back to Aristotle. For example, take what is known as *substantial change*. This sort of change involves a substance's coming into existence or ceasing to exist. There is also a sort of change known as *change of composition*. This sort of change involves replacing parts of a certain object either entirely or some part of it over time. Can a certain object then survive change of composition? The most familiar example of this is a well-known case of the *Ship of Theseus*. Suppose that we replace all of the planks of the ship one by one and out of the old planks we built a new ship. The question then becomes: do we still have the same object or ship? (For details, see Lowe 2002: Ch. 4 and Carter 1990: Ch. 6). On the other hand, contemporary metaphysicians fiercely debate over what Lewis called *the problem of temporary intrinsics*—i.e., how persisting things change their intrinsic properties. As Lewis puts it, "when I sit, I have a bent shape; when I stand, I have a straightened shape. Both shapes are temporary intrinsic properties; I have them only some of the time. How is such change possible?," (1986: 204-205; see also Hawley 2001: 16-17; Gallois 2005/2011). This is an interesting issue in its own right but for the present purposes I am not interested to pursue it in any direct way. For objection against Lewis's *temporary intrinsics* see Lowe 1988: 72-77; for Lewis's reply to Lowe, see Lewis 1988: 65-72. However, recently serious doubt has been raised as to whether there is even any puzzle about change (see e.g., Rychter 2009: 7-22).

However, we need to keep in mind that not all objects persist through qualitative change in the same way owing to the fact that different sorts of objects have different persistence conditions which are determined by their own criterion of identity (see e.g., Lowe 1989a: Ch. 2; Carruthers 1986: 72-76). In light of this, if one allows for the interrelation between persistence, change and identity to hold, then one may have some reservations, at least initially, to subscribe oneself to Heraclitus's doctrine of change, according to which, 'everything is in a constant flux like a river,' (Kirk, Raven and Schofield 1957, 1983: Ch. VI). Similarly, it is hard to endorse Parmenides's doctrine of change, according to which, 'what is [i.e., a particular object] cannot change at any time or in any respect from what it is at present' (see *Ibid*. 251-252). Heraclitus's claim above amounts to *the denial of the permanence of objects through qualitative change*; whereas Parmenides's claim amounts to *the denial of objects changing in any respect whatsoever*. Unfortunately, each of these remarks in their own way flies in the face of our experience.

Of course, someone who embraces Heraclitus's doctrine of impermanence may object to my remarks above by making a distinction between how things really are and how things appear. For example, if we drop a stick inside the pool full of water, we visually experience it as bent or twisted. Yet we know that a stick inside the pool is neither bent nor twisted. Rather it is straight. As we look up to the blue sky on a cloudless day, we experience the sun as moving. But in the post-Copernican era, we know that the Sun does not move but it is the Earth that moves on its axis. Similarly, in the case of the famous Müller-Lyer illusion, the two horizontal lines look to us as if they differ in length. Yet in reality the lines are of equal length. More can be said on this issue. But for now, the above examples suffice (see e.g., Kosso 1998: Chs. 1-2; cf. Bradley, 1908). So in light of such considerations, a defender of Heraclitus's doctrine of impermanence (by drawing some parallel) may say that our persistence over time is no different than what we see in the case of the above examples. So our sense of our personal persistence over time does not reflect the way things actually are. For two main reasons I reject such a parallel as well as the conclusion. First, the way we come to know objects in the external world is significantly different from the way we know

ourselves. For example, the former is primarily based on observation whereas the latter is primarily based on introspection. Moreover, in the case of the latter, our sense of persistence over time is rooted in our own self-consciousness, self-awareness or self-knowledge. None of these things, at least in any direct way seems to apply to the former. If I am right here, then the examples I pointed out above are disanalogus and hence fail to do the kind of work that a defender of Heraclitus's doctrine hopes that they do. Second, if as Heraclitean claims that we do not persist over time, then a number of counterintuitive consequences ensue. It seems pointless to lock up criminals for their actions, since by the time the police catch someone; a person they arrest is no longer the same person. So what is the point in arresting him? Similarly, it seems wrong to recognize people for their hard work, say for example, Peter Higgs, for his outstanding work on the prediction and subsequently on the discovery of the higgs boson/particle. If we take Heraclitus's doctrine at face value, then it remains unclear how we can avoid such highly controversial assumptions, to say the least. The sense in which we can grant Heraclitus's assumption is, only if, we take the sense of *change* in its qualitative sense as opposed to in its numerical sense (see again § 1.2.2 A). Perhaps against all such considerations, the Heraclitean may still insist that we do not persist over time. In that case, how do we know that we are numerically changing from one moment to another? Of course, here it can equally be asked: how do we know that we are not changing numerically from one moment to another? I do not belabor to settle these questions here. But in the course of my discussion in this thesis, I will give reasons why the Heraclitean objection remains to be unpersuasive.

Most philosophers who debate about the metaphysics of persistence, however, do not dispute the fact that objects persist through qualitative change (see e.g., Haslanger and Kurtz, eds., 2006: 1-26). Instead as Haslanger and Kurtz explain, philosophers' contention over persisting objects has to do with how objects persist through change (*Ibid*.). In this regard, we have three main accounts of the metaphysics of persistence. The first account is perdurantism, according to which objects persist over time by having temporal parts (Lewis 1989: 202-204; see for details Hawley 2001). The second account is, exdurantism (stage theory), according to which ordinary objects such as selves/persons are momentary stages and only persist for a period of time in which those stages exist. That is, such objects do not

last more than an instant. But according to one of its notable defenders, Sider, the stage view includes a counterpart theory of *de re* temporal predication. That is, as Sider remarks, "the truth condition of an utterance of 'Ted was once a boy' is this: there exists some person stage X prior to the time of utterance, such that X is a boy, and X bears the temporal counterpart relation to Ted," (2001: 193; also see Sider in Haslanger and Kurtz: Ch. 5). Finally, the third account is, endurantism, according to which ordinary objects persist by being wholly present throughout their lives (see for details, Merricks 1994; Haslanger and Kurtz 2006; Hawley 2001: Chs. 1-2).[50] As we shall see, my own sympathies lie with the enduratists' account of diachronic persistence.

So from the hitherto brief analysis, we can see that the three notions, namely persistence, change and identity are interrelated in such a way that they presuppose each other. But notice that at the heart of these notions lies the notion of an object. That is, we can only make sense of notions such as persistence, change and identity, if there is (in the first place) something that persists through change while retaining its numerical identity. Similarly, Simons remarks, "for something to change, it must exist before, during, and after the change, and so must survive it," (Simons in Oderberg 1999: 24). The implication here being that the notion of an object is fundamental in that notions such as persistence, change and identity are analyzed in terms of it, but the converse does not hold. This is a crucial point that has to be seen in relation to the answer given in §§ 1.1.2.3 and 1.1.2.4 to *the substantial-self/person question* (i.e., the self/person is a substance). In light of this, throughout this thesis, the question I will be exploring is this: what does it take for an ordinary familiar object like a self/person to persist through change while retaining its identity (i.e., numerical)?

C. The Difference
Returning to the complex views and the simple view, for the most part, the contemporary

---

[50] Concerning various accounts of time and persistence, see again Haslanger and Roxanne Marie Kurtz, eds., (2006: 1-26; Hawley 2001: 30-34). The three main accounts of time are: *eternalism* (all time exists, i.e., past, present and future), *presentism* (only the present time exists) and *growing block view* (only past and present time exist but no future). For an extensive reference on issues related to persistence (see Hawley 2010).

philosophical debates on this issue has been dominated by two main approaches. These are: (i) psychological criterion (e.g., Johnston 1987; Perry 1972; Parfit 1971; 1984; Shoemaker 1970; 1984; Garret 1998; Noonan 2003); and (ii) physical criterion (e.g., Williams 1956-7; Thomason 1987; Olson 1997; Ayers 1990; van Inwagen 1990).[51] On the other hand, the simple view has not been given the same level of attention in the literature (see e.g., Olson 2010). But recently effort has been made to meet this gap.[52]

But what precisely is the gist of the difference between the complex views and the simple view? As briefly explained earlier, proponents of the complex views maintain that personal identity consists in some further facts other than identity itself. By contrast, the simple view theorists claim that personal identity does not consist in any further fact other than identity itself. Defenders of the complex views also argue that, an informative, non-trivial, non-circular, non-identity involving analysis can be given for diachronic personal identity. In this regard, some form of psychological and/or physical criterion is what these philosophers have in mind. By contrast, the simple view theorists claim that no informative, non-trivial, non-circular, non-identity involving analysis can be given for diachronic personal identity. The simple view theorists therefore consider psychological or physical continuity criterion as providing us with evidential basis for personal identity as opposed to providing us with a constitutive criterion (see e.g., Lowe in Gasser and Stefan: Ch. 7; see also Lowe 1996; Swinburne 1984; Merricks 1998).

Moreover, proponents of the complex views and proponents of the simple view disagree on whether or not diachronic personal identity is different from the identity of other kinds of familiar things such as a ship, plant and so on. According to defenders of the complex views, personal identity is not different from the identity of other things (e.g., Shoemaker in Gasser and Stefan 2012: Ch. 6; see also Madell 1981: Ch.1). By contrast, proponents of the simple view argue that personal identity is different from the identity of other things and

---

[51] Each of these approaches comes in various forms.

[52] For example, in 2012 an anthology has been published by Cambridge University Press entitled *'Personal Identity: Complex or Simple?'* According to its editors Gasser and Stefan, the book's main aim is raising the profile of the simple view (2012: 1).

thus, no constitutive criterion can be given for it. But in the case of the identity of other things (e.g., trees), like the proponents of the complex views, the simple view theorists too claim that constitutive criterion can be established (see Noonan in Gasser and Stefan 2012: Ch. 4: 85-87).

But in talking about whether or not identity is different in the case of persons on the one hand and in the case of other things on the other, we are not in any way compromising our previous position on the univocality of identity across the board. That is, there are not different kinds of identity for different kinds of objects but only different situations under which we can track the identity of various sorts of objects (see § 1.2.2 A). Keeping this point in mind, the disagreement between the complex views and the simple view focuses first and foremost on whether or not a necessary and sufficient condition can be established for diachronic personal identity. For now, I have said enough by way of summarizing the central differences that exist between defenders of the complex views and defenders of the simple view. I will discuss the details involved in these issues in chapters four and five.

D. *Methodology*

As we shall see, despite over forty years of effort (i.e., since the early 1970s), the dispute besetting the question of whether or not constitutive criterion can be given for diachronic personal identity remains to be without any solution. Even those who claim that constitutive criterion can be given for diachronic personal identity, still vehemently disagree with each other as to the sort of conditions that are needed to establish it. In this regard, the debates are somewhat mired in stalemate. But why is that? I will argue that fanciful puzzle cases (often invoked in personal identity discussion) are mainly (if not solely), responsible for such a stalemate. It has become almost an established orthodoxy in personal identity literature to invoke puzzle cases as a starting point of personal identity discussions. But why should we stick to such a trend as a matter of principle? Should we not follow Wilke's (1988) lead in discussing personal identity without thought experiments? Must thought experiments put methodological constraint on how we think about the problem of personal identity? However we go about answering these and similar other related questions, I will reject a currently dominant trend that puts puzzle cases in the

driver's seat, in the sense of letting them set the agenda for discussions on personal identity What is the alternative then? My own answer will be to say that we should begin our discussion with a focus on the *Aristotelian question*.

E. Thesis Outline

This thesis is divided into four parts. Part one deals with scene setting. In the present chapter, I have done a broad survey of some of the core issues besetting the contemporary philosophical debate on the nature of the self as well as personal identity. Part two concerns with historical legacies. I will begin this part (chapter two) with the exposition and discussion of John Locke's theory of substance, the self and personal identity. In chapter three, I will discuss David Hume's theory of the self, substance, and personal identity. My discussion in both of these chapters will focus on the negative as well as the positive legacies we have inherited from these thinkers, which have significant bearing on the contemporary debate on personal identity. The discussions I advance in these two chapters will prepare us to better understand the dynamics of the contemporary debate on the nature of the self and personal identity. That said, one of the reasons to talk about Locke and Hume is to draw a contrast between their approach and that of Aristotle's.

Part three focuses on the contemporary debates on personal identity. I will begin this section (chapter four) by discussing the methodology of thought experiments. I will criticize the way thought experiments are being utilized in contemporary personal identity debates. I will relate this discussion in chapter five to the Psychological Criterion view of personal identity. Although this is not the only view of personal identity, in this thesis it will be my primary focus. This is mainly because currently this view is the most influential. So my discussion of other views of personal identity will only take place in relation to this view. In chapter six, I will present two representative views, one from a strict materialist camp (e.g., Frank Jackson's strict physicalism) and the other from a dualist camp (e.g., Lowe's non-Cartesian substance dualism) and argue that the materialistic view lacks the resources to account for a range of issues that are central to understanding the nature of the self, namely substantiality, mental causation, emergent properties, agency, self-knowledge/reflexivity, etc.

In part four, I introduce and defend (chapter seven) my own novel view which I call the *Neo-Aristotelian View of the Substantial Self/Person*. I put forth a view of a person/self that is spelled out entirely within the framework of Aristotelian substance ontology. But in doing so, everything I will say in defense of my view of the self/person will be constrained by the following conditional statement: If the Aristotelian substance ontology is true, then…. At the heart of the theory I defend, is the rejection of any sort of bundle/relational view of the self. Finally, in chapter eight, I consider some objections against the view of the self I defend in chapter seven.

I will end my thesis with a brief summary of the entire discussion presented and its outcome. My concluding remarks will point out some of the future areas of research. In this regard, my focus will be on the emergence of the phenomenal consciousness and the self.

Part II: Historical Legacies

Chapter Two

John Locke on Substance, the Substantial-Self/Person Question and Personal Identity

This chapter has two parts. The first part deals with Locke's theory of substance and the second part deals with Locke's theory of personal identity. Initially the relevance of discussing Locke's theory of substance may not be obvious. But as we shall see, Locke's theory of substance informs every major issue Locke raises in his discussion of personal identity. Unfortunately, when philosophers discuss Locke's view of personal identity, they advance their discussion without giving attention to Locke's theory of substance. Furthermore, whenever most philosophers talk about Locke's theory of substance, they often merely mention it to justify their own negative attitude toward substance ontology. In this regard, Locke's own infamous characterization of substance as 'something we know not what' is to blame. But then Locke has, as he sees it, a good reason for characterizing substance in such a way. In the first part, I want to clarify Locke's intentions in this regard thereby also arguing why we should take Locke as a realist about substance ontology. Once we do that we also need to understand how his theory of substance informs his theory of personal identity. My discussion in this regard is novel in this chapter. Since I intend to point out some serious deficiencies often underemphasized in Locke's theory of personal identity, my discussion in § 2.2 will be expositional in nature. My aim in doing this is to explain Locke's view of personal identity as closely as possible to how Locke himself understood it. In light of my discussion of Locke's theory of substance and personal identity, in § 2.2 C I will make some critical observations with respect to Locke's view of the self/person. Ultimately, this chapter will pave the way for my critical engagement with contemporary neo-Lockean views in chapter five.

It would hardly be an overstatement to say that Locke is singularly responsible for setting off the tone for the contemporary philosophical debate on personal identity. As Noonan memorably remarks in his book on *Personal Identity*, "It has been said that all subsequent philosophy consists merely of footnotes to Plato. On this topic, at least, it can be truly said that all subsequent writing has consisted merely of footnotes to Locke," (2nd ed., 2003: 24).

Furthermore, Noonan rightly observed that the majority of philosophers who write on personal identity call themselves 'Lockean' or 'neo-Lockean.' Even those who do not consider themselves as such, still develop their views of personal identity in opposition to Locke. So either way, Locke's discussion of personal identity continuous to be influential (*Ibid*). While all of this is true of Locke's continued legacy, contemporary philosophers' engagement with Locke's theory of personal identity has been and still is unduly selective. In this regard, I have in mind four aspects of Locke's view that have received much attention in the literature. These are: (i) Locke's diachronic criterion of personal identity; (ii) Locke's definition of a self/person; (iii) Locke's thought experiment, i.e., the cobbler and the prince or its modern variations; and (iv) the famous Reid-Butler objections (e.g., circularity/contradiction) against Locke's consciousness/memory criterion for personal identity over time and the subsequent Neo-Lockeans' attempt to respond to it (see chapter five).

So Noonan's remarks above in regards to Locke's place in contemporary personal identity debate, *inter alia*, are constitutive of (i)-(iv). But the same conclusion could hardly be true of Locke's substance ontology on the one hand, and what I shall call his *theological motivation* in proposing his theory of personal identity on the other. Here by 'theological motivation,' I have in mind, among other things, a Christian view of the future 'bodily resurrection' and the 'eschatological judgment,' that will ensue. These issues are severely neglected by contemporary personal identity theorists, thereby giving us a misleading picture that such issues are not central to Locke's theory of personal identity. So in this chapter, my main aim is to show, *inter alia*, that Locke is a realist about substance ontology and such ontology is central to his view of personal identity. As we shall see, Locke's own theory of substance comes in two main varieties, namely *substratum* and *particular sorts of substances*.

However, the conception of substance ontology I shall seek to adopt in defending the substantiality of the self is Aristotelian in its orientation as opposed to Lockean (see chapter one in §§ 1.1 and 1.1.2.3). (Although despite his departure from the Aristotelian tradition, the roots of Locke's own conception of substance ontology can be traced back to Aristotle). In this chapter, I will also argue that to properly understand Locke's view of personal

identity in his own terms, not taking into account his theological motivation will not be an option. This is, for at least two good reasons: (i) theological motivation is the main context within which Locke develops his view of personal identity; and (ii) the conditions Locke proposes for diachronic personal identity are directly rooted in his theological motivation (see e.g., Forstrom 2010: esp., Ch. 1). One of my main goals in chapter is to show the positive role substance ontology plays for personal identity despite Locke's own ambivalent attitude towards it.

2.1 Substratum before Locke

The notion of 'substratum,' has been around since the time of Aristotle. The notion was first introduced within the context of change, i.e., *qualitative* as well as *substantial*.[53] Following Aristotle's lead, medieval scholastic philosophers (e.g., William Ockham, Johns Duns Scotus), also extensively discussed substratum in relation to change (see for details Pasnau 2011: Ch. 2). At the heart of such discussion lies, *inter alia*, what Robert Pasnau describes as the *substratum thesis*, according to which, "all natural change requires a substratum that endures through the change," (*Ibid*. 18). So before Locke, the notion of substratum was spelled out as something that persists through change.[54] However, unlike the scholastic philosophers before him, Locke links substratum to properties as opposed to change.

2.1.1 The Need for Substratum

Locke's proposal of the theory of substratum aims at answering one key question. That is, do properties (e.g., colour) or qualities need bearers? As we shall see, Locke answers this question affirmatively. But how does Locke go about the task of accounting for the bearer of properties? How does he describe it? Does the supposed property bearer have its own nature, i.e., its own identity via which we come to know what it is? Does Locke have a

_____

[53] As we recall from chapter one in § 1.2.2 B, the first kind of change is an *alteration* in a substance (e.g., an orange changing its colour from green to yellow) whereas the second kind of change happens to a substance (e.g., when something is born or dies).

[54] Of course, the notion of things enduring via qualitative change may not be that difficult to make sense of (see e.g., chapter one § 1.2.2 B). But what do we make of the idea of something persisting or enduring via substantial change? This is the question that triggers the Aristotelian notion of *prime matter* that is supposed to be conserved through all change. Details aside, for scholastic philosophers, this is where the discussion that besets the notion of substratum proves to be extremely difficult to pin down (see for details Pasnau 2011: Chs. 2 and 3).

uniform and uncontentious way to characterize his theory of substratum? Such questions are still hotly debated and disagreements over what constitutes the right answer to them are far from over.[55]

What is a *substance*? Locke answers this question by pointing out the role of senses and reflections (*Essay* Book II, chap. XXIII).[56] Locke claims that the mind is furnished with a number of simple ideas or qualities, which are found in 'exterior things.' But what is the source of these simple ideas? Locke claims that such simple ideas are obtained via the senses. Furthermore, for Locke, the mind on its own operations is capable of noticing the unity of qualities, i.e., such qualities 'go constantly together' in experience (*Essay* II, XXIII. 1). So such unity observed in simple ideas is often attributed as belonging to one thing, even describing it under one name. But for Locke, what is being understood as one simple idea is rather a combination of many ideas (*Ibid*.). However, here we need to keep in mind that even if Locke considers an idea as subjective mental phenomenon (e.g., see *Essay* II, VIII. 8), his use of the term 'idea' is not always fixed. That is, Locke also uses the term 'idea' to refer to a quality of a subject existing external to the mind which produces a particular idea in our mind (see further Lowe 1995: 19-22).

Locke tells us that the notion of the unity we observe in simple ideas forces us to ask a question. That is, what enables these simple ideas, i.e., qualities of a physical object to stay in unity? More precisely, what underlies such unity? Initially, Locke's answer for this question may come across both as arbitrary and ad hoc. For example, Locke claims that simply because we cannot make sense of how qualities or simple ideas can subsist by themselves, we tend to assume that something grounds them or supports them. As Locke states:

> ...not imagining how these ideas can subsist by themselves, we accustom our selves, to suppose some substratum, wherein they do subsist, and from which they do result, which therefore we call substance (*Essay* II, XXIII. 1).

---

[55] I will briefly say something about the contemporary controversies in § 2.1.4. My discussion in § 2.1.1-2.1.4 has appeared in *Metaphysics or Modernity*, Baumgartner, Heisenger and Krebs, eds., 2013: Ch.1.

[56] Unless indicated otherwise, all references of the *Essay* are from the Peter H. Nidditch edition (1975).

Here some take Locke's remarks at face value and think that Locke's proposal of substratum as the bearer of sensible qualities is just a place holder. That is, its significance lies only in helping one make sense of the underlying ground for the unity of qualities whether or not the idea of substance itself is real. Commenting on the above quote, Alexander Campbell Fraser suggests:

> The expressions 'not imagining how'—'we accustom ourselves to suppose' seem to refer to our idea of substance to 'imagination' and 'custom,' instead of finding it implied in the very intelligibility of experience; for although 'custom' may explain our reference of such and 'simple ideas' or qualities to such and such particular substances, it does not show the need in reason for substantiating them, in order to conceive that they are concrete realities (in Fraser, v. 1: 390, footnote 3).

If Locke's skepticism (or seeming agnosticism) towards the reality of substance is substantiated, then it would pose a problem for Locke. Lowe remarks that, among other things, the reason why Locke's account of the idea of substance generated more controversies than any other topic we find in Locke's *Essay* has to do with its implications for theological thought. This was most importantly in relation to the accounts of God's nature and the immortality of the soul.

Doctrines such as (1) transubstantiation, i.e., a view that bread and wine transforms into the body and blood of Christ; and (2) the Trinity, i.e., the Father, Son and Holy Spirit while distinct persons, yet share the same indivisible divine nature, are rooted in the idea of substance (Lowe, 2005: 59). So the only plausible way that seems available to maintain the intelligibility of the idea of substance, as Lowe remarks, "would be to declare it innate" (*Ibid*.). In light of this, Lowe further remarks that for the religious establishments of Locke's time, Locke's empirical based understanding of the idea of substance was seen as a slippery slope down the road of atheism (*Ibid*.). This is because, *inter alia*, since God is taken to be a substance, to endorse Locke's account of substratum would require us to confess ignorance about our knowledge of God himself. But such confession (for the religious establishment of Locke's day) of our ignorance of the knowledge of God is nothing short of moving in the direction to embrace atheism.

But is the above fear of the dangerousness of Locke's view of substance justifiable? Details

aside, one way to answer this question is to look at Locke's response to one of his main critics, the Bishop of Worcester Edward Stillingfleet. As we recall, Locke employs some suspicious sounding phrases when he talks about substratum or substance: "…we accustom ourselves to suppose…" For Stillingfleet, such phrases came across as unacceptable on the basis of their implications for theological thought as briefly discussed earlier. So, in his third Letter to Stillingfleet, Locke clarifies his use of the phrase 'supposing' claiming that it should not be taken as a ground to label him as being skeptical of the reality of substance. In his Letter to Stillingfleet, Locke makes it clear that since we cannot conceive of the existence of qualities per se without being substantiated, it follows that there must be something we call substance that underlies them (see Fraser, footnotes *Essay* II, XXIII. 1: 390-391). In this case, Locke is claiming to be a realist about substance ontology. Taken this way, the suspicion we put forth earlier as to whether or not Locke is a realist about substance ontology seems to lack any ground. Thus, we can say that the notion of 'substance' for Locke is not just a place holder after all. So I disagree with Armstrong when he says that Locke's substratum, or to use Armstrong's own term 'thin particular,' is a mere postulate (1989: 60).

2.1.2 What is Locke's Substratum?

However, Locke claims that we have no idea of the notion of pure substance in general, or substratum. All that we can say with respect to substratum is that it is something that supports the qualities that produce simple ideas in us, yet we can give no further analysis of it. For Locke, pure substance is simply something 'one knows not what supports'. As Locke states:

> If any one should be asked, what is the subject wherein Colour or Weight inheres, he would have nothing to say, but the solid extended parts: And if he were demanded, what is it, that Solidity and Extension adhere in, he would not be in a much better case, than the Indian…who, saying that the world was supported by a great Elephant, was asked, what the Elephant rested on; to which his answer was, a great Tortoise: But being again pressed to know what gave support to the broad-back'd Tortoise, replied, something, he knew not what (*Essay* II, XXIII. 2; also cf. Bk. II. Ch. Xiii § 19).

Here Locke's point is that inquiry into what exactly substratum is, is not an open ended one. Even if we know that qualities (e.g., colour, weight) have bearers, they have no role to play by way of revealing the 'identity' of the thing that underlies them. So 'substratum', despite

its key role in underlying qualities, is epistemically inaccessible to us. Hence, we cannot keep on asking endlessly what substratum is. In light of this, Locke suggests that the best way to end our curiosity, to get to the bottom of the identity of substratum, is by confessing ignorance.[57] Perhaps, here Locke's emphasis on ignorance could be taken as a deterrent to unnecessary explanatory regress. Despite such *prima facie* benefit, Locke's own insistence on the unknowability of pure substance turns out to be less illuminating. But insofar as Locke is concerned, he sums up his theory of substratum as follows:

> The idea then we have, to which we give the general name Substance, being nothing, but the supposed, but unknown support of those Qualities, we find existing, which we imagine cannot subsist, sine re substante, without something to support them, we call that Support Substantia; which, according to the true import of the Word, is in plain English, standing under, or upholding (*Essay* II, XXIII. 2).

Regardless of Locke's realism about substance ontology, as we shall see, his characterization of substratum as briefly stated above did not sit well with many philosophers. But to have a complete picture of Locke's view of substratum, first we should link it to his other notion of substance, to which I now turn.

2.1.3 Locke's Particular Substance

Unlike the 'general idea of substance,' which Locke takes to be obscure and for which no positive account is available, he is much more comfortable in talking about particular sorts of substances (Essay II, XXIII, 3: 392-3). According to Locke, we come to know particular substances via combinations of simple ideas. As we saw in § 2.1.1., Locke thinks that the way simple ideas get combined or unified is through our senses, which make qualities exist together. Locke also claims that simple ideas "…flow from the particular internal constitution, or unknown essence of that substance," (*Essay* II, XXIII, 3: 393; see also John Yolton 1985: 104). Although it is unknown, according to Locke, "essence may be taken for the very being of anything, where by it is what it is," (*Essay* III, III. 15: 417). Locke claims that essence is in substances' unknown constitution of things, upon which their discoverable

---

[57] That said, however, it is important to keep in mind that for Locke, the general idea of substance is not made of via a process of combining many simple ideas, which eventually lead to the formation of complex ideas. By contrast, the abstract or general idea of substance is formed only via the process of abstraction and hence it refers to a mental process (see e.g., *Essay*, II, XI. 9). See Berkeley's objection to Locke's doctrine of abstraction in the *Introduction to the Principles* (1998).

qualities depend (*Ibid*; cf. Yolton 1985: 103-104). Locke takes such construal of 'essence' to be the proper signification of the term (*Ibid*). Here, by 'proper signification,' Locke seems to be referring to its Aristotelian sense (Lowe, 2011: 4). For example, in *Metaphysics* Z.4, Aristotle claims that essence is one of the marks by which we determine substance. The English term 'essence' is derived from the Latin word *essentia* which is the translation of the Greek phrase '*to ti en einai*' or 'what it was to be'. The other equivalent phrase Aristotle uses that captures the notion of essence is '*to ti esti*' or 'the what it is'. Aristotle tells us that the essence of each thing is what it is said to be in virtue of itself. As Aristotle puts, "For being you is not being musical; for you are not musical in virtue of yourself. What, then, you are in virtue of yourself is your essence (*Met*. Z 4 1029b 13-15). It seems then that Locke is echoing a conception of essence that takes its cue from the Aristotelian tradition.

That said, here and in other passages in the *Essay* where Locke talks about 'internal constitution' or 'essence,' he has in mind the microstructural atomic organization of a macroscopic object, which he describes as the '*real essence*.' Locke distinguishes 'real essence' from another notion of essence which he calls, 'nominal essence,' according to which the abstract general idea constitutes any particular sort of substance.[58] Although the essence of a substance (e.g., that which makes a substance to be what it is) remains to be unknown, qualities have a bearer, i.e., substance.[59] In light of this, Locke claims that we

---

[58] Locke explains abstract general idea as follows, "The use of Words then being to stand as outward Marks of our internal Ideas, and those Ideas being taken from particular things, if every particular Idea that we take in, should have a distinct Name, Names must be endless. To prevent this, the Mind makes the particular Ideas, received from particular Objects, to become general; which is done by considering them as they are in the Mind such Appearances, separate from all other Existences, and the circumstances of real Existence, as Time, Place, or any other concomitant Ideas. This is called *ABSTRACTION*, whereby Ideas taken from particular Beings, become general Representatives of all of the same kind; and their Names general Names, applicable to whatever exists conformable to such abstract Ideas….Thus the same Colour being observed to day [sic] in Chalk or Snow, which the Mind yesterday received from Milk, it considers that Appearance alone, makes it a representative of all of that kind; and having given it the name Whiteness, it by that sound signifies the same quality wheresoever to be imagin'd or met with; and thus Universals, whether Ideas or Terms, are made," (*Essay*, II, XI. 9). For controversies on Locke's notion of essence, see Atherton and Bolton in Chappell (1998: 199-213 and 214-225 res.). For further discussion on Locke's two essences (see Lowe 2005: 78-81; 1995: 76ff).

[59] Given the science of Locke's day, the ultimate constituents of microscopic objects were not known. That is why Locke confesses ignorance about his knowledge of the real essence. Hence, following a philosopher Pierre Gassendi and a scientist Robert Boyle, Locke adopted a hypothesis of corpuscularianism, according to which matter is composed of minute particles. See for details, Anstey (2000); and Barger (1976: 133-206).

come to have the ideas of a man, horse, gold, water, etc., as particular substances (*Essay* II, XXIII. 3: 393). In the traditional Aristotelian sense, things such as a particular man, a particular horse, etc., are concrete individual objects that persist through change. Although here Locke may be seen as echoing a conception of particular substances similar to that of the Aristotelian, he is using the notion of 'particular substance' in a somewhat looser sense (see e.g., Lowe 2005: 61). That is, for Locke, particular substances amount to certain ideas co-existing in unity which constitute them, i.e., they are rooted in observable qualities which make up the complex ideas of those substances.

Given such considerations, Locke rejects the scholastic notion of 'substantial form,' as having any relevance in terms of helping us to understand the nature of substances (*Essay* II, XXIII. 3: 393). Here 'substantial form,' stands for the real and immaterial principle in substances that confers them the attribute via which they are defined (see Fraser footnote 3, *Essay* II, XXIII. 3: 393). For Locke, our ideas of complex substances always have an 'obscure' and 'confused idea of 'something,' in which they inhere. That means that Locke's substratum is not identical with a particular substance, whose properties it (substratum) supports. As Locke states:

> When we talk of any particular sort of corporeal substances, as horse, stone, &c., though the idea we have of either of them be but the complication or collection of those several simple ideas of sensible qualities, which we used to find united in the thing called horse or stone; yet, because we cannot conceive how they should subsist alone, nor one in another, we suppose them existing in and supported by some common subject; which support we denote by the same substance, though it be certain we have no clear or distinct idea of that thing we suppose a support (*Essay* II, XXIII. 4: 395; also see II, XXIII. 6: 396).[60]

Although Locke calls such things as man, horse, sun, etc., particular substances, he does not consider them to be ontologically fundamental substances, which is contrary to the Aristotelian tradition (see e.g., *Cat.*, 5). Instead for Locke only three things are ontologically fundamental substances. These things are: (1) God who is eternal; (2) finite spirits or souls

---

[60] Locke makes similar remarks with respect to the operations of the mind. He says that since we cannot conceive mental states (e.g., thinking), existing without their bearer, we posit a different sort of substance which Locke calls spirit. But we have no clear and distinct idea of the substance of the spirit (*Essay* II, XXIII. 5: 395-6).

and (3) indivisible particle of matter (*Essay* II, XXVII. 2: 440). Thus, for Locke, "All other things being but modes or relations ultimately terminated in substances," (*Essay* II, XXVII. 2: 441). What is Locke's point here? The answer to his question comes from Locke's conception of what a 'mode' is.

In this regard, as Lowe explains (2005: 60-64; see also Ayers 1991: Part I), the term 'mode' which Locke uses has its root in the scholastic metaphysical tradition. But despite such terms as 'substance', 'attribute' and 'mode' being central for scholastic metaphysicians, not everyone embraced them during Locke's time. For instance, Locke himself was critical of the scholastic tradition. But for Locke, the notion of 'mode' plays a pertinent role, *inter alia*, in relation to his view of particular substances. Given the scholastic tradition, attributes are universal properties that all particular substances that fall under the same general kind have or share in common. On the other hand, modes are the particular ways those universal properties are instantiated in different particular substances. For example, a particular material substance such as a 'body' manifests the property of being spatially extended. The body also has a shape, which is the mode of the property of its spatial extension. But as Lowe further remarks, for Locke 'a particular shape' is itself a particular as opposed to a universal. That means that a particular shape of the body belongs only to that particular body. But other particular substances may still possess their own exactly similar shapes. Here note that for Locke, properties are *particulars* (i.e., tropes) as opposed to *universals*, i.e., multiply exemplifiable entities (Cf. *Essay* III, III. 1: 409).

So even if we consider the shape of a table as a 'real being,' a particular shape is not to be taken as a particular substance per se. That is, the shape of a particular table simply refers to the way a table is spatially extended. Since a 'mode' is not in and of itself a substance, it follows that a 'mode' depends both for its existence as well as its identity on a substance whose mode it is. As Lowe remarks, Locke's particular substances can be summed up as follows:

> According to Locke's stricter way of talking, it seems, many of the middle-sized material beings that we are apt to classify as particular substances—notably highly complex individuals such as plants, animals—are properly speaking only dependent entities, arising from the various complex ways in which the general attributes of

matter are instantiated by large aggregates of material atoms. That is to say, they are strictly speaking only modes, or combinations of modes (*Ibid.* 64).

2.1.4 *The Snapshot of Reactions*

As indicated earlier, even if we have good reasons to believe that Locke was a realist about substances, his characterization of substratum as '*we know not what property bearer*' understandably has brought against him a battery of criticisms from philosophers. Furthermore, as we shall see, Locke's characterization of substratum paved the way for the current prevalent anti-substance ontology attitude in contemporary metaphysics of personal identity. Although the controversies surrounding Locke's theory of substratum are important in their own right, engaging with them would take us too far afield. So in what follows, I will say just enough to give an example of some of these controversies with an aim to show their place in contemporary debates besetting the role of substance in personal identity.

As we saw in § 2.1.1, the earliest objection against Locke's theory of substance was motivated by Stillingfleet's theological worry about the Christian doctrine of the trinity. Others in Locke's time raised their objections against Locke's substratum largely based on metaphysical and epistemological grounds. In the case of the former, for example, Henry Lee echoing Aristotelian tradition, argued that the relation that exists between substance and property is asymmetrical. That is, substance is always a bearer of property but the converse is not true (see Lee 1702, II.23. 1: 110). Yet for Lee, in spite of the asymmetric relation that exists between substance and property, Locke made a serious mistake in divesting substance of its properties (II.23. 3: 110-111). So Lee's main worry against Locke's notion of pure substance has to do with the divorce Locke introduced between pure substance and property. That is, Lee's point is that Locke deprived substratum of any property. As we recall, for Locke, substratum, though it is a property bearer, is not itself knowable. But Lee argued that once we divest *pure substance* of all its properties (via which we come to know what pure substratum is), the result is to find ourselves in the dark.

Along similar lines with that of Lee's, Gottfried Wilhelm Leibniz in his *New Essays,* expresses serious doubts against Locke's characterization of pure substance. Leibniz also thinks that

divesting the concept of substance of its attributes is problematic, since the only way we come to grasp the very nature of substance is via its qualities or properties (for details see, *New Essays* II.23. 2: 218). On the other hand, unlike Lee and Leibniz, each of whose objections is metaphysical in nature, John Sergeant (another of Locke's contemporaries) raised an epistemological objection against Locke's substratum. As Sergeant remarks:

> While I perused Mr. Locke's 23th Chapter, of the Idea of Substance, I was heartily grieved to see the greatest Wits, for want of True Logick, and thro' their not lighting on the right way of Philosophizing, lay Grounds for Scepticism, to the utter Subversion of all Science; and this, not designedly, but with a good Intention, and out of their Sincerity and Care not to affirm more than they know. He fancies that the Knowledge of Substance and Extension are absolutely Unattainable….(1697, 13.22-23: 238).

Sergeant certainly gives credit to Locke for exercising an epistemological modesty by not pretending to know more than he does about the nature of substance. Yet for Sergeant, such epistemological modesty comes with a price. That is, suspending judgment in matters such as substance creates a fertile ground for skepticism. Hence, the ultimate consequence of such skepticism shakes up the very foundation of our knowledge of things. Locke's contemporaries then are unified in their insistence on the inadequacy of Locke's characterization of substratum or pure substance in general as 'something we know not what.'[61]

Beyond Locke's own immediate contemporaries, George Berkeley and David Hume also attacked the doctrine of substratum. Berkeley's objection is rooted in his doctrine of immaterialism, according to which only spiritual reality exists. Since Berkeley's immaterialism is based on the principle of '*esse est percipi*, i.e., to be is to be perceived, Locke's characterization of substratum as 'something we know not what property bearer,' stands in sharp conflict with it (see for details *Principles* § 3, 6 and 16-17). Hume, whose ontology does not accommodate a distinct property bearer, ruled out Locke's substratum as nothing more than a piece of metaphor (*Treatise*, Book I: 16; see also Lowe 1995: 83). I shall say more on Hume's view in chapter three.

---

[61] For detailed discussion of the objections of Locke's contemporaries against the doctrine of substratum, see Guta in Baumgartner, Heisenberg and Krebs (2013: Ch. 1); see also McCann (2001: 87-105).

The question remains: where does all this leave Locke's theory of substance? Is Locke's theory of substance misguided? I personally think not. As Sergeant rightly noticed, Locke's insistence on the unknowability of substratum is motivated by his epistemological modesty (i.e., by not pretending to know more than he does). Yet Locke's main goal in proposing substratum does seem to be primarily metaphysical in nature. That is, Locke on purely ontological grounds, seems to have realized that properties or qualities necessarily need some sort of bearer for them to exist. Hence, there cannot be free floating qualities (cf. Lowe 1995: 76-77). More importantly, as Lowe remarks, when Locke characterizes substratum as 'something we know not what', he (Locke) seems to be implying that it may have a nature which may be known to other beings such as angels and God (e.g. *Essay* II, XXIII. 6; Letter to the Bishop of Worcester, p. 28 as quoted in Lowe, 2000: 507). If so, the objections Locke's immediate critics raised against his theory of substance are not insurmountable.

In light of this and similar other considerations, contemporary philosophers such as C.B. Martin, E.J. Lowe, M.R. Ayers, Jonathan Bennett, J.L. Mackie, Margaret Atherton, and Martha Brandt Bolton have proposed various solutions to provide a defensible framework for Locke's theory of substance. Though the amendments these philosophers suggest differ from each other, they all agree that Locke's theory of substance is not a result of careless conjecture and thus must not be immediately dismissed. Again looking at these solutions would take us too far afield. So I will not discuss them (see, for details, Martin, 1980; Lowe, 2000 & 2005; Ayers, in Tipton, I.C., 1977; Mackie, 1976; Bennett, Ayers, Atherton and Bolton, in Chappell 1998).

One thing we can say from the hitherto discussion of Locke's view of substance is this: maintaining the distinction between the category of substance and property, clearly echoes the Aristotelian tradition.[62] Moreover, as we shall see, Locke's theory of personal identity was significantly shaped by his view of substance. However, Locke by rejecting substance as

---

[62] Also for Aristotle's influence on Descartes's, on Spinoza's and on Leibniz's conception of substance, see Woolhouse's excellent discussion (1993).

a ground of personal identity, ultimately leaves us with an unsatisfactory view of diachronic personal identity. These issues will occupy the final part of this chapter. I now turn to that discussion.

2.2 Personal Identity and the Substantial-Self/Person Question

Locke introduced both his general theory of identity and the theory of personal identity in the second edition of his *Essay*, Book II, ch. XXVII, entitled '*Of Identity and Diversity*.'[63] Since in this section my aim is to explore the role of substance ontology in Locke's theory of personal identity, I will not directly discuss Locke's general theory of identity.[64] In the context of the present discussion, Locke's relational theory of personal identity is often seen as a terminus point of the pre-Lockean era of a substance based account (see Kolak and Martin 1993: 164; Martin and Barresi 2000: 1-11). In relational theory, the 'relations,' are said to obtain among impermanent conscious memories (Kolak and Martin 1993: 164). As we shall see in chapter five, Locke's relational view of personal identity paved the way for most contemporary personal identity theorists to abandon substance ontology. Yet unlike contemporary neo-Lockeans, as we shall see shortly, Locke himself maintains both a substance ontology as well as the relational view of personal identity.

─────────────────────────

[63] Locke included this chapter in the second edition of his *Essay* at the request of his friend William Molyneux (see e.g. Locke's response to Molyneux, August 23, 1693, and March 8, 1695; see also Nidditch, 1975: xix-xxv).

[64] At the heart of Locke's general theory of identity, we find two closely related issues: (a) the *principium individuationis* (principle of individuation); and (b) sortal terms and their criterion of identity. In this section, Locke tells us how we get our ideas of identity and diversity. In this case, we get our ideas from our mind making regular comparison of the very being of a thing (e.g., organism) existing at any given time and place with itself existing at a different time and place, which enables us to form the ideas of identity and diversity (*Essay* II, XXVII. § 1: 328; cf. *Ibid*. 5-10; also Forstrom 2010: 18; Ayers 1991: part III, esp., 205-253). I shall call this, *Locke's diachronic conception of the origin of identity* (OI). This can be schematized as follows: (OI): If X exists at t_1 at a place P_1 and again if the same (i.e., numerically identical) X exists at a different time, t_2 at a place P_2, $\rightarrow$ X at t_2 = X at t_1. On the other hand, Locke's view of diversity is the notion that two things of the same kind cannot exist in the same place at the same time (*Essay* II, XXVII. § 1. 15: 328). So whenever we have distinct objects of the same kind (e.g., two cats), Locke claims that we have diversity, since objects of the same kind necessarily exclude each other thereby making co-occupation of a single region of space impossible. In light of this, Locke concludes, "one thing cannot have two beginnings of Existence, nor two things one beginning, it being impossible for two things of the same kind, to be or exist in the same instant, in the very same place; or one and the same thing in different places. That therefore that had one beginning is the same thing, and that which had a different beginning in time and place from that, is not the same but divers," (*Ibid*. 20-25). I shall call this, *Locke's exclusion principle for synchronic existence of objects* (EO). This can be schematized as follows: (EO): If X exists at t_1 at a place P_1, then it is not possible for another object of the

A*. The Order*

In chapter one in § 1.2.2 E., I suggested that our inquiry into the problem of personal identity must begin with what I called *the Aristotelian question*, i.e., 'what is X?' This is because, it is only after we tackle the 'what is X' question that we can be better prepared to take up the subsequent 'what is the identity of X' question. Although as we shall see, the 'what is X' question is inextricably linked with the 'what is the identity of X' question, it is always the former question that must take the driver's seat. In this regard, Locke begins his discussion in the right order. That is, before attempting to settle the question of wherein personal identity consists (i.e., what is the identity of X?), Locke answers the 'what is X'— i.e., 'what is a person' question. As Locke remarks,

> This being premised to find wherein *personal identity* consists, we must consider what *Person* stands for; which, I think, is a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places; which it does only by that consciousness, which is inseparable from thinking (*Essay* II, XXVII. 9: 335).

Locke's requirement for diachronic personal identity is rooted in a person's consciousness of his/her past deeds. But before we unpack Locke's idea here, first it is crucial to get clear

---

same kind X^* to be at t_1 at a place P_1. Hence, $X \neq X^*$ (see *Ibid.*; also cf. Swinburne 1968/1981: Ch.1 esp., 13ff). Locke claims that since God, Finite Intelligences and Bodies are different kinds, they do not exclude each other from the same place (see *Essay* II, XXVII. § 2ff). But these things exclude the being of the same kind. In light of such considerations, Locke concludes: "From what has been said, 'tis easy to discover, what is so much enquired after, the principium individuationis, and that 'tis plain is Existence it self, which determines a Being of any sort to a particular time and place incommunicable to two Beings of the same kind," (*Ibid*. § 3). On the other hand, Locke's treatment of (b) above is related to his discussion of the identity of 'mass of matter' and 'living organisms,' which according to Locke have different identity criterion. For example, in the case of a mass of matter, its identity consists in the co-existence of each part of the mass that makes the whole. So any addition or subtraction of the part will compromise the identity of the whole (see *Essay* II, XXVII. 3: 330). Here Bennett and Alston describe Locke as a 'mereological essentialist' (1988: 28). But in the case of living creatures, Locke maintains that their identity consists in the same life they partake as opposed to in the changing material particles that contribute to the overall internal organization of that life (see *Essay* II, XXVII. 4: 330-331). Likewise, Locke argues that the identity of animal (*Essay* II, XXVII. 5. 331) and the identity of man (*Essay* II, XXVII. 6. 331-331) consist in the life sustained under the organization of successively fleeting particles. Since Locke associates different criterion of identity to different sortal terms, there is a debate as to whether or not Locke is a sortal relative identity theorist. As Garrett puts it, "Sortal relative identity arises where objects X and Y are held to be numerically identical qua Fs, but numerically distinct qua Gs," (1998: 24). Here 'F' and 'G' stand for sortal concepts of a kind such as dog, man, etc. For a defense of relative identity, see Geach (1980). But there is no conclusive exegetical evidence in Locke's *Essay* that shows that Locke is a relative identity theorist (see e.g., Chappell, 1989: 69-83). That said, my own view is that everything Locke says in regards to the identity of a person/self, man and other things Locke presupposes absolute numerical identity (also cf. Lowe 1995: 100-101). For critical discussion of Locke's general theory of identity (see Mackie 1976: 140-145; also see, Ayer Part III: 205-253).

on three interrelated notions in the above passage. Here I have in mind, 'a thinking being', 'thinking' and 'consciousness' respectively.

1. A Thinking Being

At first glance, Locke's 'thinking being' appears to comfortably coincide with that of Descartes's *res cogitans*. But there is an important difference between Descartes's and Locke's appropriation of this phrase. As we saw in chapter one § 1.1.1.2, for Descartes, *res cogitans* is an immaterial substance. Although Locke too speaks of a person as a thinking being, unlike Descartes, he leaves the nature of a thinking being entirely unspecified (cf. Butler in Perry 1975: 101). That is, Locke does not tell us whether, in his view, a thinking being is an immaterial substance or a material substance. The main reason for this has to do with Locke's confession of our ignorance of the nature of an underlying substratum (see in this chapter §§ 2.1.1, 2.1.2 and 2.1.3; also cf. *Essay* II, XXVII. 17: 341).[65]

2. Thinking

For Locke, the noun 'thinking' covers the whole range of acts of awareness (Yolton 1985: 29). As Locke remarks:

> It [is]…impossible for any one to perceive, without perceiving, that he does perceive. When we see, hear, smell, taste, feel, meditate, or will any thing, we know that we do so. Thus it is always as to our present Sensations and Perceptions: And by this every one is to himself, that which he calls self: It not being considered in this case, whether the same self be continued in the same, or diverse Substances (*Essay* II.XXVII. 9: 335).

Here the gist of Locke's point is that when I perceive in any of these modes such as seeing, hearing, believing, etc., I also know that I am seeing, hearing, believing, etc. That is, I am aware that it is I who am perceiving (Yolton 1985: 30). So, for Locke, it is the act of reflective awareness that accompanies all modes of awareness, which in turn gives one his/her sense of self. In this case, Locke's account of how we come to have a sense of self seems to stand in sharp conflict with Dennett's claim that, "Each normal individual of this species [i.e.,

---

[65] In personal conversation, Lowe also brought to my attention that here Locke is attempting to give a neutral view of personhood that can be embraced both by dualists as well as materialists (physicalists). For Locke's assessment of Descartes's view of the self (see Forstrom 2010: Ch. 2).

Homo sapiens] makes a self. Out of its brain it spins a web of words and deeds, and, like the other creatures, it doesn't have to know what it's doing; it just does it," (1991: 416). Here Dennett's point seems to be that we get our sense of self from an involuntary activity of the brain. That is, we are just passive recipients of whatever trick our brain plays in creating a sense of a self (i.e., for each of us individually). But how could this be? If Dennett's remarks are examined from one's own first person perspective, they appear to be unconvincing. As I will argue throughout this thesis, first person awareness of oneself is hardly created as a result of one's brain playing some sort of trick, as Dennett mistakenly supposes. So following Locke, I too, maintain that our sense of self seems to be fundamentally rooted in our awareness of our own selves.

3. *Consciousness*

Locke's concept of 'thinking' as briefly spelled out above is directly related to his concept of consciousness. Although we find no systematic analysis of consciousness in Locke's *Essay*, in Book II, chapter one, Locke defines it as follows: "Consciousness is the perception of what passes in a Man's own mind," (*Essay* II. I. 19: 115).[66] What this comes down to, as Noonan remarks, is that for Locke, when one is conscious to oneself, such knowledge is only available to oneself alone. Only a person who is conscious to himself/herself is a witness to his/her own acts (2003: 42). So, if we take consciousness in the sense of 'consciousness to oneself,'[67] then for Locke, it stands for knowledge of oneself or knowledge of one's own thoughts and actions (*Ibid*. 42; see also Yolton 1993: 51). Hence, for Locke, there can be no ownerless acts of thinking, willing, believing, seeing, etc., (Yolton 1993: 51). In this regard, Locke stands in sharp contrast to Hume, who reduces everything to ownerless distinct ideas/perceptions (see e.g., *Treatise*, 1-7 & IV.VI: 251ff). In light of this, we can conclude that, for Locke, consciousness is reflexive (cf. Martin and Barresi 2000: 14).

Having briefly looked at Locke's notion of a 'thinking being,' 'thinking,' and 'consciousness,'

---

[66] For a recent defense of this conception of consciousness by Locke (see Coventry and Kriegel 2008: 221-242).

[67] Noonan attributes the locution 'consciousness to oneself' to Locke (see, 2003: 42).

we are now better prepared to look at Locke's main account of the role of substance in his account of personal identity. More specifically, my focus will be on Locke's answer to *the substantial-self/person question* and its relation to substance. I will situate this discussion within the context of Locke's theological motivations (i.e., eschatological resurrection and judgment).

B. *The Substantial-Self/Person Question, Migrating Selves/Persons and Substance*

What exactly is the self/person according to Locke's account: is it a substance or some sort of property? It is here that we begin to see Locke's interesting but yet at the same time conflicting answers that he gives to *the substantial-self/person question*. To make sense of Locke's point here, it is extremely important first to understand what I shall call the *scenario of migrant selves/persons*. For Locke, a self/person migrates from one substance to another depending on whether or not its consciousness extends to its past actions/deeds (cf. *Essay* II. XXVII. 10-11: 335-337). What does all this mean? Let us begin with the following quote by Locke:

> Whether if the same Substance, which thinks, be changed, it can be the same Person, or remaining the same, it can be different Persons (*Essay* II. XXVII. 12: 337).

Here Locke's question has two parts. The first part concerns whether, despite a change in substance, there can be the same person; the second part concerns whether, despite being the same substance, there can be different persons. In responding to the first part of the question, Locke claims that it is difficult to determine the sameness of immaterial substance, since we cannot be certain regarding whether the consciousness of past actions can be transferred from one thinking substance to another. This is even more difficult, as Locke seems to think, in the face of the possibility of "one intellectual substance representing to it, as done by it self, what it never did, and was perhaps done by some other Agent" (*Ibid*. 13: 338). On the other hand, regarding the second part of the question (i.e., which concerns whether despite being the same substance there can be different persons), we may ask: what possible scenarios could we think to answer this question? If we let Locke himself answer this question, here is what he has to say:

> Whether the same immaterial Being, being conscious of the Actions of its past Duration, may be wholly stripp'd of all the consciousness of its past Existence, and

lose it beyond the power of ever retrieving again: And so as it were beginning a new Account from a new Period, have a consciousness that cannot reach beyond this new State (*Essay* II, XXVII.14: 338).

Locke's point here is that for the same immaterial being to be a sustainer of two distinct persons, two conditions have to be satisfied. Condition 1 (C1) requires the loss of the substance's entire past memories/recollections of its deeds (beyond any hope of restoration). Here we need to keep in mind that, for Locke, memory is first personal or biographical, which presupposes one's remembering of past action as one's own (see Lowe 2000: 277-283).[68] Condition 2 (C2): requires the substance to have an entirely new beginning of experience that has no link whatsoever with the previous one. So for Locke, if (C1) is satisfied, then that would bring about 'the loss of the earlier person'; whereas if (C2) is satisfied, then 'a new person' emerges. To make sense of this, we only need to understand Locke's famous criterion of memory, according to which a person *P* is identical with himself/herself of a year ago or two weeks ago, etc., if he/she remembers the thoughts he/she had or the action he/she had done a week ago or a year ago, etc. More specifically, following the Lowe-Olson-Hudson schemata (see chapter one §1.2.1) for the 'criterion for identity,' we can say that for Locke, if a person P_2 at t_2 and a person P_1 at t_1 are the same, then P_2 is identical with P_1 if and only if P_2 and P_1 satisfy the memory condition as described above. But what if the memory condition is not satisfied, i.e., in this case, if P_2 fails to remember P_1's past thoughts or actions? For Locke that would mean that P_2 is not identical with P_1 and vice versa. So for Locke, having an active memory of one's past thoughts or action is decisive for one's identity over time.[69]

In light of this, Locke thinks that given (C1)-(C2), we can have a scenario where the same substance can be said to be the sustainer of two distinct persons. But Locke's remarks here

---

[68] Here Locke's notion of memory is first personal or biographical memory. Other variants of memory include factual memory (e.g., I can be said to remember that George W. Bush was the President of the USA prior to President Obama) and memory of practical skills (e.g., I can be said how to ride a bike (see further Robert Audi 1998: Ch. 2).

[69] Of course, here we should not overlook what constitutes a person's identity at a time as well. But in the case of personal identity at a time, we cannot appeal to memory. Here it has often been said that multiple personality syndrome as well as split-brain syndrome raises a challenge in our effort to identity a person with itself at a time. I will have something to say about this issue in chapter eight.

might bring some odd, and in fact, ultimately an implausible picture to our mind regarding human persons. For example, though the analogy is not perfect, let us take the old fashioned cassette tapes or any modern day CDs and other electronic recording devices. The songs recorded in these recording devices can be entirely wiped out and be replaced with entirely new music/songs. Suppose that we remove the songs from one of the CDs or cassette tapes and record on it entirely new songs. In this case, we have the same CD or cassette tape but distinct songs. From this, it follows that the CD or cassette tape can be recycled numerous times. But such a model, even if it makes perfect sense in the case of cassette tapes, seems to be less than helpful to understand what constitutes personhood in the case of human persons, to say the least. But I won't defend my claim just yet, since my goal here is to look at Locke's view on its own ground.

Locke goes on defending the plausibility of (C1)-(C2) by appealing to those who believe in the pre-existence of the soul. As Locke states:

> All those who hold pre-existence, are evidently of this Mind, since they allow the Soul to have no remaining consciousness of what it did in that pre-existent State, either wholly separate from Body, or informing any other Body; and if they should not, 'its plain Experience would be against them. So that personal Identity reaching no farther than consciousness reaches, a pre-existent Spirit not having continued so many Ages in a state of Silence, must needs make different Persons (*Essay* II, XXVII.14: 338-339).

Locke, following what he already stated in (C1)-(C2), shows us here how the pre-existence of the soul with no prior remaining consciousness can be a sustainer of different persons. To bring his point home, Locke remarks:

> Suppose a Christian *Platonist* or *Pythagorean*, should upon God's having ended all his Works of Creation the Seventh Day, think his Soul hath existed ever since; and should imagine it has revolved in several Human Bodies, as I once met with one, who was perswaded his had been the Soul of Socrates…how reasonably I will not dispute (*Essay* II, XXVII.14: 339).

Since for Locke, only consciousness constitutes personal identity, it (consciousness) can be sustained by different souls/substances. So Locke asks: "…would any one say, that he, being not conscious of any of Socrates's Actions or Thoughts, could be the same Person with Socrates?" (*Ibid*.). Here Locke expects us to say that he who does not remember the actions

of Socrates cannot be the same person as Socrates. Locke realizes that an immaterial spirit is taken as something that thinks in us and also keeps us the same through the constant change of our bodies. But in the absence of the consciousness of past Actions, Locke rules out any hope of having the same person even if we still have the same soul. In light of this, Locke says:

> But he, now having no consciousness of any of the Actions either of Nestor or Thersites, does, or can he, conceive himself the same Person with either of them? Can he be concerned in either of their Actions? Attribute them to himself, or think them his own more than the Actions of any other Man, that ever existed? So that this consciousness not reaching to any of the Actions of either of those Men, he is no more one self with either of them, than if the Soul of immaterial Spirit, that informs him, had been created, and began to inform his present Body, though it were never so true, that the same Spirit that formed Nestor's or Thersites Body, were numerically the same that now informs his. For this would no more make him the same Person with Nestor, than if some of the Particles of Matter, that were once a part of Nestor, were now a part of this Man, the same immaterial Substance without the same consciousness, no more making the same Person by being united to any Body, than the same Particle of Matter without consciousness united to any Body, makes the same Person. But let him once find himself conscious of any of the Actions of Nestor, he then finds himself the same Person with Nestor (*Essay* II, XXVII.14: 339).

1. *Context*

Here Locke's discussion on personal identity is generally considered to be rooted in three central theological convictions. These are personal immortality, bodily resurrection and the final judgment day (see e.g., Forstrom, 2010). Details aside for now, Locke's convictions about these three theological issues frame the context within which he discusses personal identity. But recently Galen Strawson, in his book entitled *Locke on Personal Identity: Consciousness and Concernment,* remarked that the basic idea behind Locke's discussion of personal identity does not depend on the Day of Judgment in any way (2011: XX-XIII). As Strawson puts it:

> The idea of the Day of Judgment is no doubt, a fantasy, but the fundamental idea behind Locke's discussion of personal identity does not depend on it in any way although he was bound to pose the question of personal identity in eschatological terms in which his conception of personal identity (or moral identity) is intuitively natural, is independent of the story of the Day of judgment (*Ibid*.).

Having set the scene by dismissing Locke's theological convictions, Strawson in the remainder of the book gives an extended defense of the forensic (i.e. moral) notion of a person. Strawson also accuses Butler, Reid and Berkeley of seriously distorting Locke's view of personal identity (see *Ibid*. Ch. 1).[70] But are Strawson's remarks right in the above passage? The answer is not at all. Contra Strawson, nothing Locke says in his *Essay* on personal identity makes sense without its theological context. In fact, an eschatological judgment and the resurrection are so central for Locke that he discusses them in great detail in his unpublished essay, "*Resurrectio Et Qua Sequuntur*" (see King 1829: 316-323). The point here is that Locke never wrote his chapter on personal identity out of mere philosophical curiosity (cf. Forstrom, Ch. 1). In light of this, the very forensic notion of a person that Strawson defends in his book, as I will argue, is inextricably linked with Locke's theological context (also cf. Noonan 2003: 24). Hence, I find Strawson's remarks above to be entirely misguided.

Once Strawson's type of mistakes are dismissed, it becomes clear that the stance Locke takes on personal immortality, bodily resurrection and the final judgment day, directly shapes his engagement with key philosophical approaches of his time, namely Cartesian dualism, Hobbesian materialism, Cambridge Platonism and Boyles' corpuscularian mechanism (see further Forstrom, 2010: Ch. 1). In light of this, as I already mentioned, whatever Locke says on personal identity is aimed at making sense of the above three theological convictions he firmly endorses.

With this brief background in mind, we can see why Locke continues to stress the centrality of one's being conscious of one's past actions and thoughts to ensure the sameness of the person. Once we grant this, Locke thinks that we may be able to conceive the person at the resurrection and the person prior to the resurrection to be the same even though the body of the resurrected person will not be the same as the body the person had prior to that resurrection. Locke drives his point home with his most quoted thought experiment:

---

[70] Here it is not my interest to discuss Strawson's criticism of Butler, Reid and Berkeley. My point here to show that Strawson himself cannot escape a similar criticism he attributes to others.

For should the Soul of a prince, carrying with it the consciousness of the Prince's past Life, enter and inform the Body of a Cobbler as soon as deserted by his own Soul, every one sees, he would be the same Person with the Prince, accountable only for the Prince's Actions: But who would say it was the same Man? (*Essay* II, XXVII.15: 340).

The point of Locke's thought experiment here is that a person goes wherever his/her consciousness goes—this is an insight that has had immense influence on the contemporary neo-Lockean philosophers (e.g., see Noonan, 2003). For Locke, a Cobbler with the Prince's consciousness should be considered to be the same person as the Prince. Notice here that Locke introduces the notion of accountability into his consideration of why a Cobbler with the Prince's consciousness is only responsible for the actions of the Prince. As we shall see, for Locke, the last Judgment Day is linked with accountability which, in turn, is central to his notion of a 'moral person'.

But in light of Locke's above thought experiment, the question remains: Are the terms 'person' and 'man' coextensive? Locke realizes that given the ordinary way of speaking, 'person,' and 'man' are co-extensive in that they stand for one and the same thing. This is because, as Locke sees it, everyone has a liberty to speak as well as apply certain ideas to objects as he/she thinks fit. For Locke, one also has a liberty to change the ideas one associates with objects as often as one pleases. However, Locke quickly makes remarks that indicate that we should not expect co-extensiveness with respect to spirit [substance], man, and person. As Locke puts it:

But yet when we will inquire, what makes the same Spirit, Man, or Person, we must fix the Ideas of Spirit, Man, or Person, in our Minds; and having resolved with our selves what we mean by them, it will not be hard to determine, in either of them, or the like, when it is the same, and when not (*Essay* II, XXVII.15: 340).[71]

But why should we think that a difference in ideas can be a basis, as Locke seems to think, to deny the co-extensiveness of substance, man and person? For example, triangle and trilateral are different concepts but the terms are coextensive. Why can't we likewise treat

_____

[71] This is one of those passages where Locke introduces what Noonan calls a 'tripartite ontology, namely person, man and substance' (2003).

Locke's tripartite terms, i.e., substance, man and person as coextensive? Locke would not want to treat these terms as co-extensive. One of the reasons has to do with the place Locke gives for consciousness, since that is what constitutes personal identity. As Locke puts it:

> Had I the same consciousness, that I saw the Ark and Noah's Flood, as that I saw an overflowing of the Thames last Winter, or as that I write now, I could no more doubt that I, that write this now, that saw the Thames overflow'd last Winter, and that view'd the Flood at the general Deluge, was the same self, place that self in what Substance you please, than that I that write this am the same my self now whilst I write (whether I consist of all the same Substance, material or immaterial, or no) that I was Yesterday. For as to this point of being the same self, it matters not whether this present self be made up of the same or other Substances, I being as much concern'd, and as justly accountable for any action was done a thousand Years since, appropriated to me now by this self-consciousness, as I am, for which I did the last moment (*Essay* II, XXVII.16: 340-341).

Once again this passage shows, as do many of the others we have seen so far, Locke's emphasis on the centrality of consciousness in ensuring the sameness of the self on the one hand and the centrality of substance as its bearer on the other (regardless of substance contributing nothing at all in grounding personal identity). So for Locke, insofar as the self is conscious of its own past actions, the question of personal identity can be settled. But what precisely does Locke mean by a self? Locke tells us that a self is a conscious thinking thing regardless of whether it is made up of spiritual, or material simple, or a combination of both. Moreover, a self as Locke describes it is sensible, or conscious of pleasure and pain. A self is also capable of happiness or misery. A self is also concern'd for itself as far as that consciousness extends (*Essay* II, XXVII.17: 341).

In light of this, Locke wants to link the notion of the self and personal identity. Given the above definition of the self, Locke thinks that we can make sense of the right and justice of reward as well as punishment; happiness and misery. As we recall, notions such as reward and punishment are rooted in eschatology or in what will happen at the last Judgment Day. So the bottom line for Locke is that personal identity consists not in the identity of substance but in the identity of consciousness. It is only then that a person stands fit before God to receive either reward or punishment for his/her actions.

But what if I lose the memory of some parts of my life beyond any hope of retrieving them. Am I not the same person who did those actions even if I am not able to recall them now? Locke responds to this objection by telling us what the first person pronoun 'I' is applied to. Locke claims that 'I' is applied here only to man. But if the same man is taken to refer to the same person, then 'I' can be taken to also refer to the same person. However, if it is possible for the same man to have two distinct incommunicable consciousnesses over a period of time, then we get two different persons (*Essay* II, XXVII. 20: 342). That seems to be why, as Locke thinks, human laws do not punish the mad man for the sober man's actions nor the sober man for the mad man's actions (*Ibid*.).

But Locke expects that his own response to the above objection may not be appealing. This is because, as Locke claims it is difficult to conceive how the same individual man can be two persons. To help us get a grip on this difficulty, Locke gives us three ways by which we can understand what is meant by the same individual man: (1) it could mean the same individual, immaterial thinking substance or the same numerical soul, and nothing else; (2) it could mean the same animal in total abstraction from an immaterial soul; and (3) it could mean the same immaterial spirit united to the same animal (*Essay* II, XXVII. 21: 343). In light of what he has already said about the centrality of consciousness being a ground for personal identity, Locke denies (1)-(3) as contributing anything substantial towards personal identity. We arrive at this conclusion, as Locke thinks, because of the difference that exists between the identity of man (*Essay* II. XXVII.6: 331-332) and person (see details, *Essay* II, XXVII. 21: 343).

For Locke, as we have been seeing all along, the emphasis on consciousness as the only ground for personal identity is motivated by his conviction about the last Judgment Day. Locke tells us that in the great [Judgment] Day, God will reveal every hidden secret in the human heart. But what justifies anyone's facing punishment is a person's being conscious of his/her deeds. This is the only condition that needs to be satisfied to be able to answer to God. Therefore, as Locke's thinks, no one will be responsible for what he/she cannot remember (*Essay* II, XXVII. 22: 343-344). Here Locke re-echoes his emphasis on the centrality of consciousness, which is what unites remote existence into the same person. In

short, for Locke only consciousness makes self. That means that, as Locke further claims, self is not determined by identity or diversity of substance; rather it is determined only by identity of consciousness (*Essay* II, XXVII. 23: 344). So for Locke, a substance can only be part of the self insofar as the consciousness of the self is attached to it, otherwise it is no longer part of what one calls one's self (*Essay* II, XXVII. 24: 345).

Unlike Parfit (1984), who argues that what matters in survival is not identity, Locke seems to be stressing the opposite point. Locke claims that every intelligent being sensible of happiness or misery must grant that there is something that is himself/herself that he/she is concerned for (*Essay* II, XXVII. 25: 345-346). As we shall see, for Locke, at the center of concern for oneself lies one's own future identity, which makes a difference during the future resurrection as well as the Final Judgment Day. If so, what then grounds the continuity of the self over time as the same self? Before we answer this question, we need to keep in mind that, in some sense, what Locke says about the non-centrality of the bearer of consciousness (i.e., substance) in grounding personal identity lends some support for Parfit's claim that what matters in survival is one's counterpart's existence who is only psychologically connected with the prior self. There are lots of issues we can raise at this point; but for now we won't pursue that discussion until chapter five. That said, contra Parfit, we can safely say that Locke does not rule out the importance of the concern one has for one's own future identity. In light of this, Locke remarks:

> This self has existed in a continued Duration more than one instant, and therefore 'tis possible may exist, as it has done, Months and Years to come, without any certain bounds to be set to its duration; and may be the same self, by the same consciousness, continued on for the future. And thus, by this consciousness, he finds himself to be the same self which did such or such an Action some Years since, by which he comes to be happy or miserable now (*Essay* II, XXVII. 25: 345-346).

So for Locke the continuity of the self over time is grounded in the continuity of the same consciousness. But Locke further points out the bearer of consciousness itself does not have to be the same numerical substance. Several substances over a period of time may have been united or separated from the self. But what matters, as Locke sees it, is the continuity of the same self with the same consciousness, despite the diversity of its bearers. Hence, Locke concludes:

And so we have the same numerical Substance become a part of two different Persons; and the same Person preserved under the change of various Substances….Any Substance vitally united to the present thinking Being, is a part of that very same self which now is: Any thing united to it by a consciousness of former Actions make also a part of the same self, which is the same both then and now (*Ibid*. 346).

I will close my expository discussion of Locke's view of personal identity with another key aspect of his theory, that is, the forensic notion of a person. I now turn to that discussion.

2. *Locke's Notion of a Moral Person*

Towards the end of his *Essay* Locke shifts his focus to the notion of a moral person. As we recall, Locke opens his *Essay* with what can fairly be described as the ontological characterization of the notion of a person according to which, a person is: (1) a thinking intelligent Being, that has reason and reflection; (2) one that can consider itself as itself; and (3) the same thinking thing in different times and places; which it does only by that consciousness, which is inseparable from thinking (*Essay*, II. XXVII. 9: 335).

Locke's notion of a moral person is situated within the context of the three theological convictions that we pointed our earlier, namely personal immortality, resurrection and the last Judgment Day. Taken in the moral sense, 'person' is a technical term. Locke also understands the term 'person' to be synonymous with the term 'self'. Taken in the technical sense, Locke says that person is a forensic term, appropriating actions and their merit. In this case, Locke tells us that the term person belongs only to intelligent agents who are capable of a law, happiness and misery (*Essay* II, XXVII. 26: 346). Locke claims that to make a person be accountable for his/her deeds, his/her personality must extend itself beyond the present existence into the past. But how does that happen? Locke tells us that it happens by consciousness, which becomes concerned and accountable, owns and imputes to itself past actions. For Locke, one's concern for happiness and any other experiences of a person are determined to be one's own, only by consciousness. As Locke states:

And therefore whatever past Actions it cannot reconcile or appropriate to that present self by consciousness, it can be no more concerned in, than if they had never been done: And to receive Pleasure or Pain; i.e. Reward or Punishment, on the account of any such Action, is all one, as to be made happy or miserable in its first

being, without any demerit at all. For supposing a Man punish'd now, for what he had done in another life, whereof he could be made to have no consciousness at all, what difference is there between that Punishment, and being created miserable? (*Essay* II, xxvii. 26: 346-347).

It is clear that for Locke, what justifies a person's receiving reward or facing punishment depends on whether or not a person appropriates past actions to himself/herself. But what does Locke's notion of 'appropriating' one's past actions amount to? In this regard, Kenneth Winkler gives us what seems to me to be an unsatisfactory analysis of the notion. According to Winkler, for Locke, "consciousness of a past act is merely a representation of it as one's own; it is not knowledge of the pre-existing fact that the act was one's own" (in Vere Chappell, 1998: 154).

Winkler's suggestion here is a bit puzzling. If appropriation of a past act is no more than a mere representation of it as one's own, then we will be forced to say that a person deserves reward or punishment for acts he/she has never done. All that a person needs to do is to just to be able to represent certain past acts as his/her own. But that seems to be deeply implausible. In fact, in light of the immediate context of Locke's discussion here, Winkler's suggestion does not reflect Locke's view. So contra to Winkler, the truth of the matter here seems to be exactly the opposite of what Winkler states in the second half of his remark, i.e. appropriation does not require knowledge of the pre-existing fact that the act was one's own. This point is made particularly clear by Locke's own remarks:

> And therefore conformable to this, the Apostle tells us, that at the Great Day, when every one shall receive according to his doings, the secrets of all Hearts shall be laid open. The Sentence shall be justified by the consciousness all Persons shall have, that they themselves in what Bodies soever they appear, or what Substances soever that consciousness adheres to, are the same, that committed those Actions, and deserve that Punishment for them (*Essay* II, xxvii. 26: 347).

So for Locke, what ultimately justifies God's distribution of reward or punishment at the last Judgment Day, is squarely contingent on what a person appropriates to himself or herself. But such appropriation is not, as Wrinkler mistakenly thinks, a mere representation of past acts as one's own; rather it is a genuine attribution of past acts as one's own with the

conviction that those acts are done by oneself.[72] That being said, our hitherto discussion on Locke's forensic notion of a person can also be captured in light of John Rawl's highly influential *Theory of Justice* (1971/1999). Central to Rawl's *theory of justice*, *inter alia*, is the notion he describes as, 'justice as fairness' (*Ibid*. 3). The basic idea behind this notion has to do with an equal treatment of persons regardless of their social status. Rawl argues that the equality of all people consists in nothing but in the fact that, "each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override....," (*Ibid*. 3). Similarly, Locke's main point is that at the last Judgment Day, in distributing punishment or reward for the actions of persons, God must exercise 'justice as fairness'.

The gist of Locke's notion of *Identity* and *Personal Identity* can be summed up as follows:
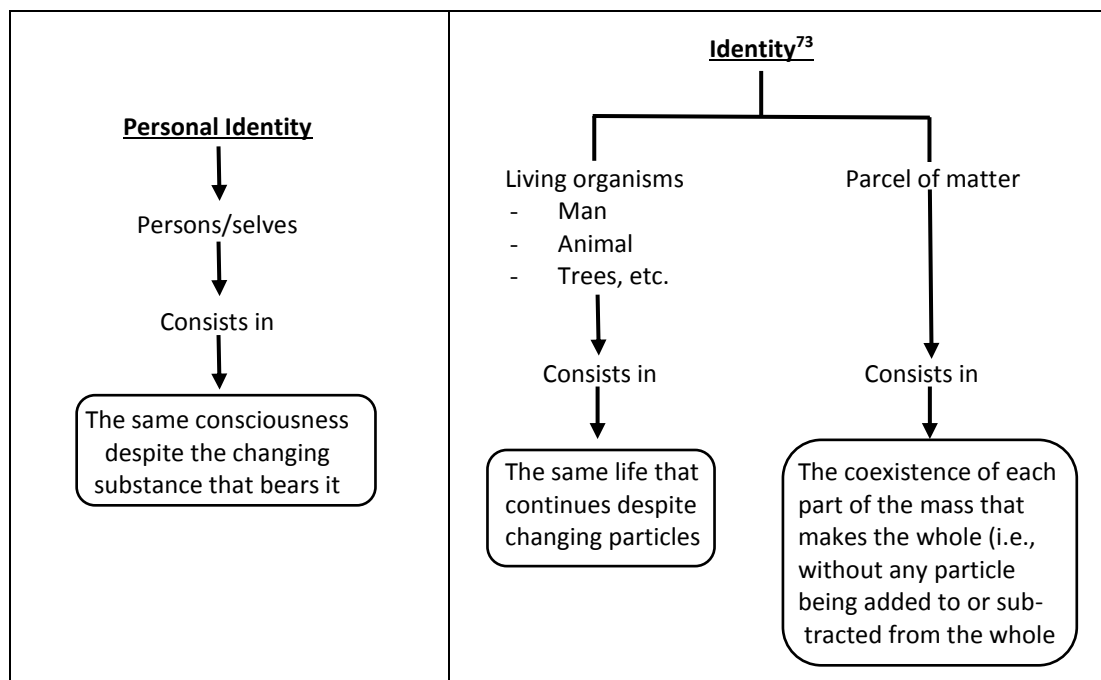
Figure 1   Figure 2

Having looked at what Locke has to say about substance, identity and personal identity, I will end this chapter with some observations I think are still unresolved in Locke's theory of

---

[72] Having covered a lot of ground regarding wherein constitutes personal identity, Locke ends his Essay by giving a brief discussion of the difficulty which we still have in understanding the nature of the substance as well as the problems that arise from misapplying names to various ideas. Locke ends the chapter by claiming that continued existence makes identity (*Essay* II, xxvii. 27-29).

[73] For a summary of Locke's general theory of identity, see again footnote # 64.

personal identity.

C. *Outstanding Issues in Locke's Theory of Personal Identity*

Again what exactly is the self/person, according to Locke's account: is it a substance or some sort of property? In light of our discussion in § 2.2 A and B as well as in § 2.1.3, Locke has left us with the following conflicting and perhaps even contradictory answers (cf. Flew 1951; Shoemaker 1963: 45-46; also cf. Chisholm 1976: 108):

1. A person is an intelligent thinking being (i.e., be it a material or immaterial substance).
2. When a person thinks, it is substance (immaterial) that thinks in a person.
3. A self/person is a conscious thinking thing.
4. There are only three fundamental substances: (i) God; (ii) finite souls; and (iii) indivisible particles of matter. Everything else dissolves into complex modes or relations.
5. Man is not a person owing to the different persistence or identity conditions each requires. For example, man's identity consists in the 'same life' whereas a person's identity in the same 'consciousness.'
6. Self/person is constituted by consciousness.

Of (1)-(6) above, (4) obviously rules out a self/person from the class of substances. So based on (4), we can say that for Locke, the self/person is not a substantial entity. But then, (3) seems to conflict with (2). This is because given (2), it is not a person that thinks after all, but an immaterial substance that thinks in a person, whereas given (3), a person is said to be an entity that thinks. How can Locke maintain both (2) and (3) at the same time? Again (1) seems to stand in direct tension with (2), (3) and (4). To see this, we need to pay attention to (1) where Locke seems to be saying that a person is a substance even though we do not know its nature, i.e., whether it is material or immaterial. But if a person is a substance in either sense, i.e., material or immaterial, then Locke faces three difficult questions: (i) what justifies Locke's exclusion of a person from the class of substances as evidenced by (4)?; (ii) what justifies Locke's denial of a person as a thinker as evidenced by (2)?; and (iii) what justifies Locke's definition of a self/person as a thinker which he has already seemed to have denied as evidenced by (2)? Finally, since Locke denies that the notion of 'man' coincides with the notion of 'person', as (5) indicates, we are left speculating what a self/person is supposed to be. So given (6), it may be thought that, since a self/person is not a substantial entity, it can only be constituted by consciousness. More

precisely, selves or persons do seem to be logical constructions out of the consciousness that annexes itself to various changing substances. To be even more precise, a self/person is a complex mode (see e.g., LoLordo 2012: Ch. 2; see also Figure 2 in § 2.1.3). If this is true, then ultimately, Locke's two key characterizations of a self/person in (1) and (3) above make no sense whatsoever.

Along similar lines, Chisholm also points out his worry about the Lockean notion of the transfer of a self/person from one substance to another. In this case, Chisholm remarks:

> But if I am placed in a certain thinking substance and am not identical with that thinking substance, then there are *two* different things—the thinking substance and I. But if there are two things, which of us does the thinking? (1976: 108).

Chisholm claims that there are only four possibilities to answer the question at the end of the above quotation (*Ibid*): (a) Neither of us does the thinking—that is, neither of us thinks. (b) I think but the thinking substance does not think. If so, Chisholm asks: why call the latter a 'thinking' substance, then? And what relation do I bear to this thinking substance given that proper parts of substances are themselves substances? But if I am myself a thinking substance, what is the point of saying there is *another* thinking substance in which I am 'placed' or to which I am 'annexed'? (c) the thinking substance thinks but I do not. If so, Chisholm claims that this is absurd. This is because, given Locke's view, it is not really I who think; rather, it is some other thing that thinks in me—some other thing that does what I mistakenly take to be my thinking'; and (d) both the thinking substance and I think. Isn't this multiplying thinkers beyond necessity? If I want my lunch, does it follow that two of us want my lunch? Or does the thinking substance want its lunch and not mine?

Chisholm's observations above entirely coincide with my own earlier observations in (1)-(6) above. I agree with Chisholm when he says it is reasonable to conclude that there is no significant sense in which we can make sense of the transfer of a self from one substance or individual thing to another (1976: 108).

At this point, we should admit that Locke's answer to *the substantial-self/person question* is a real mess, to say the least. To my knowledge, Locke has not given us any solution by way

of dealing with this mess nor did he seem to be aware of the problem himself in his theory. Although I do not intend to come to Locke's defense on this front, it is worthwhile commenting on what led him to such a problem in the first place, and the possible way out for Locke. With respect to the first point, it seems to me that the main source of the problem for Locke's answer to *the substantial-self/person question* has to do with his own unsettled view of the role of substance in his theory. Locke, largely owing to his deep commitment to his empiricist epistemology and metaphysics, denied substance as a ground for personal identity (cf. LoLordo 2012, esp., Ch. 2). But the question then is: how does consciousness fair better in grounding personal identity compared to a substance? I answer this question by way of suggesting one possible way Locke could redeem what I described above as a *mess* in his theory. Although for reasons we have already seen, Locke adamantly clung to the consciousness/memory criterion for personal identity over time, there is a sense in which a substance based account would do a better job for all the same reasons Locke opted for a memory based criterion.

That is, when a person fails to remember his/her past actions, what is happening is not as Locke mistakenly seems to think, there is an absence of the original person in that particular substance. Rather, what has happened is the failure of the manifestation of a certain disposition, which in Locke's case would be a *memory*. That means that failing to manifest a certain disposition (for whatever reason), does not show that that person's identity is compromised. Rather what it means is that the same person who endures over time is not able to manifest that particular disposition. If Locke had understood that the manifestation or the lack thereof a certain disposition would not compromise a person's identity, then he would have happily identified a self/person with a substance. In doing so, he would have given us a much more defensible and plausible theory of personal identity over time. How then would my suggestion above have eased Locke's fear about the last Day Judgment?

Here we can say that since, as theists believe, God is all-knowing and all-powerful,[74] even if

_____

[74] For present purposes, I do not concern myself with qualifying in what sense such divine attributes should be understood. My point here is simply that God's knowledge and power are such that He can know and do what seems to be impossible to know and do from the standpoint of humans.

a certain person fails to remember his past actions, God would still have ways to distribute fair reward and punishment. For example, God could reinstate the disposition a person has lost and on that basis proceed with his reward and punishment duty. It seems then that Locke's assumption that the loss of a disposition necessarily implies the migration of a person from one substance to another is unconvincing. I will return to similar issues in chapters five and eight.

Having discussed Locke's view of substance and personal identity, in the next chapter, I turn to Hume.

Chapter Three

David Hume on Substance, the Substantial-Self Question and Personal Identity

In chapter two, I attempted to show, *inter alia*, how Locke's theory of *substance*, *the substantial-self/person question* and *personal identity* are all interrelated. I argued that even if in Locke's theory, substance does not ground personal identity, it still plays an important role in terms of being a bearer of consciousness. Yet, Locke's rejection of substance as a ground for personal identity did not work out in his favour. In this case, as we saw, Locke's conflicting set of answers to *the substantial-self question* rendered his theory to be less than satisfactory. Having discussed Locke's view of personal identity and substance, it is important to turn to Hume's. One of the main reasons is that, properly understanding how Locke and Hume thought about personal identity will put us in a better position to get a complete picture of their influence on contemporary personal identity theorists. For example, Parfit can be taken both as a paradigm example of someone who represents the Humean—the reductionist view of persons and personal identity (see e.g., 1984: Part III; also cf., Behan in Tweyman III 1995: Ch. 94 and Capaldi in Tweyman III 1995: Ch. 87; Noonan 2003: 85) as well as the neo-Lockean tradition. I will discuss Parfit's own view in chapter five. So the discussions of the last chapter as well as the present chapter are meant to prepare us for chapter five, where my focus will be on the contemporary personal identity debates with reference to the neo-Lockeans.

In this chapter, I will discuss *substance*, *the substantial-self/person question* and *personal identity* from Hume's standpoint. Hume discusses these issues in Book I of *A Treatise of Human Nature* (hereafter *Treatise*).[75] As we shall see, these three notions together constitute what I shall call a *three-pronged source* for what Hume has to say, *inter alia*, about change, the persistence of objects over time, the identity of objects and what it means to be a subject of experience (i.e., a self/person). The other part in the Treatise that is relevant to understand Hume's discussion of the above issues is the *Appendix*, which is

---

[75] Hume does not discuss personal identity in his An *Enquiry Concerning Human Understanding* (EHU). All my references to Hume's *Treatise* are from L.A. Selby-Bigge's second edition (1978). For more recent editions of Hume's *Treatise* (see Norton and Norton 2000).

included at the end of Book III.[76] As is well known, in the *Appendix*, Hume openly confesses that his theory of personal identity is unsuccessful and yet he does not give us any tangible reason(s) for his recantation (cf. Penelhum 2000: 99).[77] Rather than follow the current orthodoxy in putting on the table yet another speculative reason for Hume's recantation, I will draw some positive lessons from it and show how such lessons can be used in defense of a realist account of substance ontology—which is the main focus of this thesis. So in this chapter, my goal is both to explain Hume's account of substance ontology in his own terms and also examine whether the challenges he raises against it pose any serious threat for those who defend it.

3.1 *Background Assumption*

As Locke's empiricism was based on the premise that all knowledge is rooted in sense perception and reflection (see e.g., chapter two), Hume also holds that experience is the sole window into any inquiry that we engage in. Hume strongly believes that, since the science of man is the only solid foundation for all other sciences, it (i.e., the science of man) must be rooted in experience and observation (*Treatise* XVI). Hume thinks that we cannot go beyond experience. But as Benedict Smith points out, Hume's view of experience is multi-faceted, ranging from introspection to commerce with others (unpublished paper, pp. 4-8). So for Hume, any hypothesis or philosophical system that pretends to discover the fundamental qualities of human nature must be rejected as presumptuous and chimerical (*Ibid*. XVII; see Lindsay 1964: Intro.; Smith 1941; Norton, ed. 1993).

3. 2 *Hume's Project*

Given his empiricist metaphysics and epistemology, as we shall see, in the *Treatise*, Hume's discussion of *substance, the substantial-self/person question and personal identity* is negative. However, some philosophers claim that Hume's negative discussion of these

---

[76] Unless otherwise indicated, I discuss Hume's view of the self and personal identity based on *Treatise Book I* and the *Appendix*.

[77] This has left contemporary philosophers but to speculate a great deal about the reasons for Hume's dissatisfaction with his own theory of personal identity. Unfortunately, in this regard, there is no consensus in sight (see e.g., Garrett 1981: 337-358; Swain in Traiger 2006: 141-148; Behan in Tweyman III 1995: Ch. 94; Capaldi in Tweyman III 1995: 631-334; Traiger in Tweyman III 1995: 716; Pitson 2000: Ch.4).

things is part of his strategy in clearing the way for his own positive account of the self that comes later in *Book II of the Treatise* (see e.g., Capaldi 1975: 92-93; cf. Baillie 2000: 37ff).[78]

So philosophers who take this line of argument view Hume as a realist about the self (see e.g., Bricke 1980: ch. 4; cf. Swain in Traiger 2006: ch. 8; Passmore 1976; Stroud 1977; McIntyre 1979, 1989; Pike 1967; Beauchamp 1979; Triager 1985; Flage 1990; Loeb 1992). In light of this, it has been claimed that whatever criticisms Hume advances against the self, or a substance, such criticisms only target a certain conception of them. For example, Bricke remarks,

> Hume does not deny the existence of selves. Nor does he deny *tout court* that selves are substances. What he does deny is a certain philosophical theory, the substrate theory, that purports to explain or elucidate the fact that selves are substances (see e.g., Bricke 1980: ch. 4; see also Norton 1982: IX and Introduction).

For reasons that we shall see, Bricke's as well as other like-minded philosophers' characterization of Hume's stance on the nature of the self or substance is highly questionable. I will argue in this chapter that Hume's view on s*ubstance*, *the substantial-self/person question* and *personal identity,* should be taken at face value (e.g., cf. Sybil Wolfram 1974: 586-593; Noonan 1999: Ch. 5, 2003: Ch. 4; see also Chisholm 1969: 7-21; Perry 1975: 26-30; also cf. Penelhum 2000: Chs. 3-6). Given this approach, as we shall see, it will be entirely unclear whether Hume's negative discussion of *substance*, *the substantial-self/person question* and *personal identity* is motivated by his desire to offer us his own positive alternative view of the self or a substance. So I take philosophers who think otherwise to be confusing their own reconstructed view of what they believe or think Hume is saying (or perhaps even should have said) about these things with the actual view Hume clearly endorses in the *Treatise*. To make sense of all of this, first we need to ask one central question. That is: what is Hume's project in the *Treatise* regarding *substance*, *the substantial-self/person question* and *personal identity*?

---

[78] Other books by Hume that philosophers use to establish their claim about Hume's realism about the self include: *An Inquiry Concerning Human Understanding*, see Beauchamp (1999); *An Enquiry Concerning the Principles of Morals*, see Beauchamp (1998).

I take Hume to be a philosopher who sets himself up for the task of, 'deconstruction'.[79] The straightforward dictionary sense of this term can be described as tearing down something that was built up. But for present purposes, I intend to use the term 'deconstruction' as described by Samuel Wheeler III, in his *Deconstruction as Analytic Philosophy*. According to Wheeler, "Deconstruction as Analytic Philosophy primarily addresses analytic philosophers, especially those who do not think that other traditions in philosophy might be capable of contributing something of interest to philosophy," (2000: 1). Here Wheeler's characterization of 'deconstruction,' nicely fits what Hume does in the *Treatise*. In this case, Hume's deconstructionism targets three main traditions, namely *the Aristotelian*, *the Cartesian* and *the Lockean.*

3.3 *Hume's Framework*

To put Hume's deconstructionism project in its proper context, first we must begin with his doctrine of ideas. Central to Hume's theory of ideas is the notion of perception, which he says constitutes two distinct kinds, viz., impressions and ideas. By impressions, Hume has in mind those perceptions with the most force and violence (impression of reflection). An example of such impressions include: sensations, passions and emotions. On the other hand, by ideas, Hume refers to what he describes as the faint images of such things in thinking and reasoning (*Treatise*, 1; see also *An Abstract to Treatise*, 647). For Hume, the difference between impressions and ideas consists in the degrees of force and liveliness, with which they strike the mind and make their way into our consciousness (*Ibid*.). Although there is a resemblance between impression and ideas, Hume claims that such resemblance is not universally true (*Ibid*. 3). This is because, in Hume's view, impressions are always superior to ideas in a sense that the former is the source of the latter but not vice versa (Ibid. 4-5). As Pears remarks, "Hume's empiricism is rooted in his axiom that all our ideas are derived from impressions," (1990: vii). In the literature, this is often labelled as 'the copy principle,' (see e.g., Morris and Brown 2014: § 1.4). Hume himself sums up the primacy of

---

[79] I borrowed this term from a famous literary critique Jacques Derrida (see e.g., Royle 2003: ch.3; Stocker 2006).

impressions over ideas as "the first principle [he established] in the science of human nature," (*Treatise*, 7).

Hume further expands his discussion of the primacy of impressions over ideas by distinguishing between simple and complex perceptions (cf. chapter two § 2.1.1; Locke's *Essay* II, chapters, VII-IX and XII). Simple perceptions or impressions and ideas admit of no distinction or separation whereas complex perceptions are distinguishable into parts. In the case of the first, for example, Hume claims that the idea of red which we form in the dark and the impression which strikes our eyes in bright sunlight differ only in degree not in nature (*Treatise*, 3). To illustrate the latter, Hume uses an example of a particular colour, taste and smell as qualities that are all united in a particular apple. Here Hume correctly points out that these are various qualities which can be distinguished from each other (*Treatise*, 2).

From this, Hume may be taken to be endorsing 'a two-category ontology'. That is, if the various qualities that are said to be united in a particular apple are distinguishable from each other on the one hand and the apple in which they inhere on the other, then it is sensible to conclude that an apple is a distinct bearer of those qualities. Here in the words of Heil (2012: 2-3) we have 'complementary categories': the apple is a bearer/a substance and the properties are ways that the apple is. But Hume, as we shall see, denies any of such analysis. What is his alternative proposal? The answer emerges from Hume's deconstructionism, which will be discussed shortly.[80] As we shall see in due course, the very distinction Hume draws between impressions and ideas on the basis of criterion such as 'force' and 'vivacity' is far from clear. Regardless of such problems, however, for Hume

---

[80] Hume further divides impressions into two kinds, namely sensation and those of reflection. For Hume, the former arises in the soul from unknown causes whereas the second is derived from our ideas. As Hume states: "An impression first strikes upon the senses, and makes us perceive heat or cold, thirst or hunger, pleasure or pain of some kind or other. Of this impression there is a copy taken by the mind, which remains after the impression ceases; and this we call an idea. This idea of pleasure or pain, when it returns upon the soul, produces the new impressions of desire and aversion, hope and fear, which may properly be called impressions of reflexion, because derived from it. These again are copied by the memory and imagination, and become ideas; which perhaps in their turn give rise to other impressions and ideas. So that the impressions of reflexion are only antecedent to their correspondent ideas; but posterior to those of sensations, and deriv'd from them" (*Ibid*. 7-8).

impressions and ideas are pertinent for his deconstructionism to get off the ground. However, here the key point to keep in mind is that for Hume, impressions and ideas are the primary sources of the concepts or ideas we form regarding objects of our experience (cf. Fodor 2003). Hume's view of concept acquisition does not focus on mastering the meaning of a particular word to understand the concept of what that particular word stands for. That method of concept acquisition is what the later Wittgenstein seems to suggest in his *Philosophical Investigations* (see e.g., 1953: 208). Hume's view of concept acquisition is quasi-inference based. (This is not to be confused with strict logical/deductive inference). How does Hume's quasi-inference model work? Hume explains this by introducing three interrelated notions, namely *resemblance*, *association* and *cause and effect*, each of which also plays a critical role in Hume's deconstructionism. For Hume, these notions together establish a framework within which he analyses the role of impressions and ideas in the process of concept acquisition. In Hume's own words, I will sum up these notions under § A below.

A. Resemblance, Association, and Cause and Effect

Hume claims that we all experience the connection between ideas. For example, simple ideas give rise to complex ones. But the question remains: What is responsible for the connection or association of ideas? In other words, what is the 'uniting principle' among ideas? It is mainly in response to this question that Hume proposes the three notions mentioned above. As Hume describes:

> 'Tis plain, in the course of our thinking, and in the constant revolution of our ideas to any other that resembles it, and that this quality alone is to the fancy a sufficient bond and association. 'Tis likewise evident, that as the senses, in changing their objects, are necessitated to change them regularly, and take them as they lie contiguous to each other, the imagination must by long custom acquire the same method of thinking, and run along the parts of space and time in conceiving its objects. As to the connexion….'Tis sufficient to observe, that there is no relation, which produces a stronger connexion in the fancy, and makes one idea more readily recall another, than the relation of cause and effect between their objects….These are therefore the principles of union or cohesion among our simple ideas, and in the imagination supply the place of that inseparable connexion, by which they are united in our memory (*Treatise*, 11-12; see also *EHU* §§ II and III).

One of the key questions that comes from Hume's remarks in this passage can be put as follows: Is the principle of unity a figment of our own imagination or something that we come to grasp in the nature of things that we experience? I shall call this: *The Disjunctive Question.* From Hume's standpoint, the answer to *The Disjunctive Question* mainly depends on what we take the principle of union/association of those complex ideas (that arise from simple ideas) to amount to. Hume divides such complex ideas into *Relations*, *Modes* and *Substances* (*Treatise*, 13). Of these three complex ideas, relations have to do with objects that admit of comparison. Hume analyses such '*philosophical* relations' under seven general domains (*Treatise*, 14-15). These are: (1) resemblance; (2) identity; (3) space and time; (4) all objects which admit of quantity, or number; (5) objects that possess the same quality in common; (6) contrariety and (7) causes or effects. That said, before taking up my discussion of Hume's deconstructionism of *substance*, *the substantial-self/person question* and *personal identity*, in the next section to follow, I will briefly point out some difficulties that beset Hume's distinction between impressions and ideas.

B. *What Is the Difference?*

As we recall, for Hume, the 'force and vivacity' attached to impressions are the basic features that distinguish it from the ideas. But how are we supposed to understand 'force and vivacity'? As Wayne Waxman points out, there are difficulties in understanding these notions. To see this, Waxman asks us to recall that in Hume's view, the difference between impressions and ideas rests on those features of perceptions that immediately make their way to consciousness and those we come to know via experience and custom (see again § 3.3). For example, the main principle of Hume's theory of ideas, i.e., that every idea is copied from an antecedent impression, is experiential in nature. The only elements of Hume's principle that are independent of experience (i.e., accessible to immediate perception) are the resemblance between impressions and ideas and the temporary precedence of one against the other (see e.g. *Treatise* 73 and 168f). But for Hume, even the resemblance relations we notice between impressions and ideas are the result of associative imagination (*Treatise* 10). In light of this, Waxman asks: is there any immediate perceptible difference between impressions and ideas, prior to, and independent of, both experience and associative imagination that warrants their distinction? (1994: 27).

Hume's response to this question is affirmative, since he thinks that we are aware of the 'force' and 'vivacity' among the perceptions of the mind with no recourse to experience (*Ibid*.). Given such a response, Waxman claims that it is not surprising to see Hume relying on some quality (perception) that can be directly sensed and felt (*Ibid*.). But the question remains: what quality of perception does Hume have in mind? The answer for this question is far from clear. As Waxman puts it:

> One possibility is that it [quality of perception] is something like the difference between scarlet and brick red; or between red and purple; or perhaps that between blinding light and dim. But it is equally possible that actual seeing — even in a pitch black room (i.e. uniformly black visual sensation) — and mere recollection (even that of an intensely bright, sharply delineated scene, viewed only a moment before, i.e., an image of sensation). Or perhaps in yet another direction: a directly discernible feature not of the objects present to consciousness themselves (i.e. perceptions) but of our consciousness of them (i.e. a kind of primitive intentional attitude adopted towards perceptions). It is even unclear whether impressions and ideas are not simply diverse aspects of perceptions, so that a perception counts as an impression if it is regarded as a present content of consciousness and an idea if deemed a representation of something distinct from it (*Ibid*.).

Given the lack of clarity on what Hume's notion of 'force' and 'vivacity' amount to, it remains hard to see the merit of Hume's distinction between impressions and ideas (see further *Ibid*.). Yet as we shall see, Hume's deconstructionism has much to do with this distinction. So to properly understand what Hume has to say about other issues (e.g., substance), we cannot help but pay close attention to Hume's distinction. Having put in place the relevant background information, I now turn to my discussion of Hume's deconstructionism of *substance*, *the substantial-self/person question* and *personal identity*.

3.4. *Hume's Target*

What Hume says in the *Treatise* about each of these things, i.e., *substance*, *the substantial-self/person question* and *personal identity* is interrelated. Hence, for the most part, as we shall see, the objections Hume raises against one can equally be applied against the other. This is because Hume uses similar tools to deconstruct certain conceptions of these things. Here by 'similar tools,' I mean to refer to Hume's doctrine of ideas which was discussed in § 3.3. That said, the first instance of Hume's deconstructionism against substance comes in § VI entitled '*Of modes and substances*.' Hume remarks:

I WOU'D fain ask those philosophers, who found so much of their reasonings on the distinction of substance and accident, and imagine we have clear ideas of each, whether the ideas of substance be deriv'd from the impressions of sensation or reflexion? If it be convey'd to us by our senses, I ask, which of them; and after what manner? If it be perceiv'd by the eyes, it must be a colour; if by the ears, a sound; if by the palate, a taste; and so of the other senses. But I believe none will assert, that substance is either a colour, or sound, or a taste. The idea of substance therefore be deriv'd from an impression of reflexion, if it really exist. But the impressions of reflexion resolve themselves into our passions and emotions; none of which can possibly represent a substance. We have therefore no idea of substance, distinct from that of a collection of particular qualities, nor have we any other meaning when we either talk or reason concerning it (*Treatise*, 15-16).

In this passage, Hume has raised what I take to be a three-pronged objection against substance ontology. First, Hume rejects the classical substance-attribute distinction. Let us call this: the *Distinction Objection*. In classical theory, as Armstrong explains, "It is natural to distinguish a thing, an individual, a token, from any particular properties that the thing happens to have," (1989: 59). An example of such distinction would be an apple (i.e., substance) and its redness, size, shape, etc., (i.e., its attributes). Second, Hume denies that one can have a clear idea of substance and attribute/property. Let us call this: the *Knowledge Objection*. Hume's third objection concerns with the origin of substance. Hume's point is that if substance exists, then it must be locatable in impression or reflection (see again footnote # 80). Since in Hume's view, one cannot locate substance in an impression, there will be no other way to establish a distinct category of substance. Let us call this: the *Origin Objection*. For Hume, the above three objections feed into each other in a sense that the first objection is rooted in the second one and the second one is rooted in the third one. That is, as Hume sees it, we can only be justified in drawing the distinction between substance and its attributes, if it were the case that we had a distinct idea of each. But to have a distinct idea of each (i.e., substance and accident/property), first we would have to know how substance could originate from impressions of reflection. So taken from Hume's standpoint, in attempting to respond to the above objections, one cannot isolate one—i.e., must respond to all of them. With this challenge put in place, Hume's conclusion is that we have no idea of substance distinct from a collection of qualities. So Hume further remarks:

The idea of a substance as well as that of a mode, is nothing but a collection of simple ideas, that are united by the imagination, and have a particular name assigned them, by which we are able to recall, either to ourselves or others, that

collection. But the difference betwixt these ideas consists in this, that the particular qualities, which form a substance, are commonly referred to an unknown something, in which they are supposed to inhere; or granting this fiction should not take place, are at least supposed to be closely and inseparably connected by the relations of contiguity and causation (*Treatise*, 16; see also *Ibid*. 16).

Here we can see that Hume understands Locke's substratum (i.e., an unknown something) to be a fiction, dispensing altogether with the notion of substance understood as a distinct bearer of properties. As we shall see, throughout his discussion of substance in relation to the self as well as personal identity, Hume hangs onto Locke's infamous characterization of substance as 'unknown something'. In doing so, he wants to show why it is a mistake to endorse realist substance ontology. Here by 'realist substance ontology,' I mean a conception of substance as a property bearer, as pointed out in chapters one and two. In deconstructing the asymmetrical relation that exists between a substance and the properties it bears, Hume's end goal is to establish his own bundle theory, according to which a thing is identified with the bundle of its properties. Of course, among famous British empiricists, Hume is not the first in proposing a bundle theory. Berkeley before him already argued that material objects are a collection of sensible qualities (see e.g., *Principles*, 53). But unlike Hume, regarding the mind, Berkeley was not a bundle theorist, since he maintains the distinction between the mind/self and its properties (see *Ibid*.). So compared to both Berkeley and Locke, Hume took a radical step of reducing ordinary familiar objects to a mere collection of properties. In doing so, Hume may be taken as one of the early pioneers of the tradition of a one-category ontology defended by contemporary eminent philosophers such as Russell (1940), Williams (1953), Campbell (1990), Hochberg (1964), Castañeda (1974), Ayer (1954) and Simons (1994).

3.4.1 *What is Next?*

In raising the three pronged objections as outlined above, Hume's next move is to use them collectively to deconstruct *substance*, *the substantial-self/person question* and *personal identity*. In this case, Hume's deconstructionism of these things is located within the three traditions identified in § 3.2, namely the *Lockean*, the *Aristotelian* and the *Cartesian* traditions. As we saw above, Hume already dismissed Locke's notion of substratum calling it

a fiction. Similarly, in 'Of the Antient Philosophy,' in Part IV of the *Treatise* in § III, Hume remarks:

> I am persuaded, there might be several useful discoveries made from a criticism of the fictions of the antient philosophy, concerning *substances*, and *substantial forms*, and *accidents*, and *occult qualities*; which, however unreasonable and capricious, have a very intimate connexion with the principles of human nature (*Treatise*, 219).

But what are the reasons that persuaded Hume to label the ancient philosophical conception of substance, *inter alia*, as a fiction? Hume explains:

> 'Tis evident, that as the ideas of the several distinct *successive* qualities of objects are united together by a very close relation, the mind, in looking along the succession, must be carry'd from one part of it to another by an easy transition, and will no more perceive the change, than if it contemplated the same unchangeable object. This easy transition is the effect, or rather essence of relation; and as the imagination readily takes one idea for another, where their influence on the mind is similar; hence it proceeds, that any such succession of related qualities is readily consider'd as one continu'd object, existing without any variation. The smooth and uninterrupted progress of the thought, being alike in both cases, readily deceives the mind, and makes us ascribe an identity to the changeable succession of connected qualities….In order to reconcile which contradictions the imagination is apt to feign something unknown and invisible, which it supposes to continue the same under all these variations; and this unintelligible something it calls a *substance, or original and first matter* (*Treatise*, 220; see also *Ibid*. 221-225)*.*

The gist of Hume's point comes down to answering what I earlier called the *Disjunctive Question*. That is: *Is the principle of unity a figment of our own imagination or something that we come to grasp with in the nature of things that we experience?* If Hume's above account with respect to how we come to form a belief in a continuing object is true, then we have no other option but to endorse the first disjunct of the *Disjunctive Question*. But in endorsing the first disjunct, that is in considering the *unity* we ascribe to objects of our experience as a figment of our imagination, Hume is effectively telling us to rule out the second disjunct. Notice again that the second disjunct presupposes the *unity* we experience in the object of our experience to be grounded in the nature of the things themselves, as opposed to being the result of our own psychological extrapolation. So given Hume's argument, the 'or' in the *Disjunctive Question* is to be understood as an exclusive 'or' as opposed to it being an inclusive 'or'. That is, as we know from rules of elementary logic, the disjunction, *P* v *Q*, is true in the exclusive sense of the term iff one but not both disjuncts are

true. By contrast, if the 'or' in the disjunction, *P* v *Q*, is taken in its inclusive sense, then each disjunct could be true, although we are presented with a choice to pick one of the disjuncts. Put this way, as we shall see, Hume's account of the *unity* of things, regardless of whether we are concerned with the objects we experience in the external world or those of our own mental life, is highly questionable.

Most importantly, given the *Disjunctive Question*, Hume's suggestion that we can do away with the second disjunct in favor of the first disjunct hardly holds water unless Hume gives us a compelling reason to do so. Now we can see at least why Hume is adamant in opposing the second disjunct in the *Disjunctive Question*. That is, if one embraces the second disjunct, then one cannot avoid its implication for substance as a principle of unity. What should we do then? Should we take Hume's conclusion and move on? Not without a fight, at least. Hume should earn our concession to his conclusion. If Hume establishes his case for the irrelevance of realist substance ontology, then those of us who are sympathetic to such a realist ontology cannot help but concede to him. But will Hume deliver a better option? At this point, I can only say that it is yet to be seen. In making such remarks, I am expressing my disagreement with those philosophers who seem to have ruled out that Hume's rejection or critique of the doctrine of substance is successful (see e.g., Flage 1990: Ch. 4; Allaire in Tweyman III 1995: Ch. 47).

Hume's criticism of the notion of substance goes beyond the Lockean and the Aristotelian traditions. For example, in Part IV of the *Treatise* in § V, Hume turns to Cartesianism. Here, Hume's target is the Cartesian doctrine of the immateriality, simplicity and indivisibility of a thinking substance (*Treatise,* 240-244). As he did in the case of the previous two traditions, Hume also insists that for the Cartesian doctrine of immaterial substance to have any chance of being real, it must meet certain conditions. As Hume states:

> As every idea is deriv'd from a precedent impression, had we any idea of the substance of our minds, we must also have an impression of it; which is very difficult, if not impossible, to be conceiv'd. For how can an impression represent a substance, otherwise than by resembling it? For how can an impression represent a substance, since, according to this philosophy it is not a substance, and has none of the peculiar qualities or characteristics of a substance? (*Treatise*, 232-233).

Here Hume has in mind what I had identified earlier as *the Knowledge Objection* (which requires one to have a clear idea of substance) and *the Origin Objection* (which requires an account of the origin of substance). In the remainder of § V, Hume argues that since the philosophers he criticizes have failed miserably in terms of meeting these objections, the whole idea of substance in general and Cartesian immaterial soul in particular, is not worth talking about. As Hume states:

> Thus neither by considering the first origin of ideas, nor by means of a definition are we able to arrive at any satisfactory notion of substance; which seems to me a sufficient reason for abandoning utterly that dispute concerning the materiality and immateriality of the soul, and makes me absolutely condemn even the question itself….We have no perfect idea of any thing but a perception. A substance is entirely different from a perception. We have, therefore, no idea of a substance. Inhesion in something is suppos'd to be requisite to support the existence of our perceptions. Nothing appears to support the existence of a perception (*Treatise*, 234).

Here by particular definition, Hume is referring to Descartes's own definition of substance as '*something which may exist by itself*,' (*Treatise,* 233; cf. see § 1.1.1.2 footnote # 12). Hume claims that if this definition is true, then perceptions also qualify as having the status of substances, since they can exist by themselves without needing any bearer. In saying this, Hume is once again raising the *Distinction Objection* we saw earlier. That is, there is no need for a distinct bearer of perceptions. Why? As Hume thinks, we have no idea of substance. But Hume's claim here about the actual possibility of the existence of ownerless perceptions, is a moot point to which we should pay close attention. It would be very interesting to know how it would be possible for ownerless perceptions to exist (see also *Treatise*, 222). This is an issue I will take up when I discuss Hume's account of personal identity. That said, as we shall see, Hume extends his deconstructionism to personal identity as well. Before I take up that discussion, it is worth reflecting on Hume's hitherto deconstructionism against substance ontology.

3.4.2 *Hume's Psychologism*[81]

The term '*psychologism*' is often used to show what one takes to be the mistake of

---

[81] For debate on physchologism in general as well as its original source, see Martin Kusch (2007/2011). But here in discussing *Hume's Psychologism*, I am neither assuming nor implying that psychology is irrelevant in discussing substance ontology in general or personhood in particular.

identifying non-psychological with psychological entities (Kusch 2007/2011). For example, Frege thought that it is a mistake to base mathematics upon psychology, since the latter is vague and inexact. Here Frege is running what is understood to be an anti-psychologistic argument (see e.g., 1884: 38ff; see also Kusch 2007/2011 § 4). Along similar lines, what I am here calling *Hume's Psychologism* is intended to show why Hume's account of the source of our idea of substance is problematic. Hume's account of the source of our idea of substance is psychological in nature, since as we saw, he appeals to an impression of sensation or reflection. But how well does Hume's psychologism fair if seen in light of the three-pronged objection identified earlier, namely: (a) *the Distinction Objection*; (b) *the Knowledge Objection*; and (c) *the Origin Objection*? Again here we need to evaluate the merit of these objections against the backdrop of the three traditions Hume's deconstructionism targets, namely the *Lockean*, the *Aristotelian* and the *Cartesian*.

In this case, (a) targets the distinction between substance and the properties it bears. Since Aristotle, Locke and Descartes each makes such a distinction, the objection certainly attracts their attention. But (b) that focuses on one's claim to have a clear and distinct idea of substance does seem to attract only Descartes's attention. This is because, Descartes clearly claims that he has a 'distinct and clear idea' of the distinction between immaterial substance and material substance (see e.g., *Principles of Philosophy* AT VIII 60: CSM 213; see also chapter one § 1.1.1.2). However, Descartes does not say that such an understanding of the nature of substance comes via derivation from some sort of impression. So the force of (b) against Descartes's claim will not be that strong. How about (b) against Aristotle? As we shall see in chapter seven, Aristotle begins his theory of substance by taking the reality of ordinary familiar objects like human beings for granted. That is, for Aristotle people, trees, stones, etc., are real. So Aristotle would not have been moved by (b), since he does not exercise conceptual inference to establish the reality of such things. But this does not mean that Aristotle does not use concepts to think about things. That is a different matter. I will have something to say about this issue in chapter seven.

What can we say about Locke's take on (b)? I would say that Locke would not have been moved by (b) either. As we have discussed in chapter two, for Locke, the general idea of

substance is not made via a process of combining many simple ideas, which eventually lead to the formation of complex ideas. By contrast, the abstract or general idea of substance is formed only by the process of abstraction and hence it refers to a mental process. Moreover, Locke does not claim to have a clear idea of the nature of substance at all. So Locke can easily turn his back against (b). That leaves us with Hume's (c), which deals with the origin of our idea of substance. This is the objection that Hume cannot afford to lose against Aristotle, Locke and Descartes. I say 'Hume cannot afford to lose' because Hume's *Psychologism* is rooted in (c) so much more than in (a) and (b). Whose attention here will Hume get then, that is Aristotle's, Locke's or Descartes's? Unfortunately, given our discussion of Descartes's and Aristotle's account of substance in chapter one, and also given our discussion of Locke's account of substance in chapter two, Aristotle, Descartes and Locke would all reject Hume's (c). The main reason for this is that none of these philosophers came up with their conception of substance ontology via a mechanism of Hume's sort of *psychologism*.

So Hume cannot just get what he wants by setting up the terms of the debate in such a way that only serves his case. Hume needs to take his opponents' views of substance on their own terms and give us reasons why we should reject them. As things stand, at least for now, Hume has not done that. In light of this, therefore, it is reasonable to conclude that Hume's *Psychologism* is entirely of his own making. That is to say that, Hume has not yet established that the problem he tries to locate in the views of other philosophers is something that his opponents would readily admit is actually a problem to their views. Hence, it remains unclear what significant role his *Psychologism* could play in enabling Hume to succeed in his deconstructionism project (cf. Oderberg 2012).

But Hume's failure to establish (c) against Aristotle, Locke and Descartes, significantly weakens the force of his other objections, i.e, (a) and (b). This is precisely because as I had pointed out earlier, the three-pronged objection feed into each other in a sense that (a) is rooted in (b) and (b) is rooted in (c). So what are Hume's chances to bounce back again? I will explore an answer to this question and other similar ones in Hume's discussion of personal identity. I now turn to that discussion.

3.5 *The Substantial-Self/Person Question* and *Personal Identity*

In the *Treatise* Part IV § VI entitled, '*Of Personal Identity*,' we come across Hume's official presentation of his view of the self and personal identity.[82] Hume's theory of personal identity is shaped, *inter alia*, by his criticism of substance. As we recall, the heart of Hume's criticism of substance has to do with his rejection of an independent principle of unity. We saw this point in detail in § 3.4.1 where I analyzed *the Disjunctive Question*. As we recall, the *Disjunctive Question* asks: *Is the principle of unity a figment of our own imagination or something that we come to grasp in the nature of things that we experience?* For Hume, there is no straightforward way to deal with this question. So his own solution as we saw in our earlier discussion was to deconstruct it. As we shall see, this question also crops up repeatedly in Hume's discussion of personal identity. Here too, Hume's goal is to show the irrelevance of a certain way of understanding the principle of unity.

3.5.1 *The Starting Point*

Unlike Descartes and Locke, each of whom began the analysis of the ontology of the self/person with the *Aristotelian Question* (i.e., what is X?), Hume's starting point is with the unity of the self/person. As Pears remarks, "Hume shows little direct interest in the criteria for identifying a person across an interval of time," (1990: 121; see also Noonan 1999: 189). The reason for this, as we shall see, partly has to do with Hume's own conception of identity. That said, Hume opens his discussion of personal identity by attacking the rival views. As Hume states:

> There are some philosophers, who imagine we are every moment intimately conscious of what we call our SELF; that we feel its existence and its continuance in existence; and are certain, beyond the evidence of a demonstration, both of its perfect identity and simplicity. The strongest sensation, that most violent passion, say they, instead of distracting us from this view, only fix it the more intensely, and make us consider their influence on self either by their pain or pleasure. To attempt a further proof of this were to weaken its evidence; since no proof can be deriv'd from any fact, of which we are so intimately conscious; nor is there any thing, of which we can be certain, if we doubt of this (*Treatise*, 251).

---

[82] Hume discussed the notion of the self in his other writings as well. In this regard, for an excellent discussion on Hume's unpublished works, as well as posthumously published works, see McIntyre in Tweyman III (1995: Ch. 93).

As Robison rightly points out, Hume thinks that the philosophers he has in mind [most likely Descartes being one of those] are mistaken in attributing to a self/person three essential features: (a) it continues to exist through the whole course of our lives; (b) it remains the same throughout its existence; and (c) we are intimately conscious of it (Robinson in Tweyman 1995: 690). But why does Hume reject the features in (a)-(c)? Unsurprisingly, Hume's answer is that these features fail to meet the *Origin Objection*. As we recall from our discussion in § 3.4 above, *the Origin Objection* has to do with the inability to account for the origin of substance. In this case, Hume raises *the Origin Objection* in relation to the self. As Hume remarks:

> Unluckily all these positive assertions are contrary to that very experience, which is pleaded for them, nor have we any idea of *self*, after the manner it is here explain'd. For from what impression cou'd this idea be driv'd? This question 'tis impossible to answer without a manifest contradiction and absurdity; and yet 'tis a question, which must necessarily be answer'd, if we wou'd have the idea of self pass for clear and intelligible. It must be some one impression, that gives rise to very real idea. But self or person is not any one impression, but that to which our several impressions and ideas are suppos'd to have a reference. (*Treatise*, 251).

Here Hume's point is that unless the *Origin Objection* is met, any claim one makes about the ontology of the self turns out to be unjustified. Moreover, Hume sets the bar so high for a defender of the self, when he says:

> If any impression gives rise to the idea of self, that impression must continue invariably the same, thro' the whole course of our lives; since self is suppos'd to exist after that manner. But there is no impression constant and invariable. Pain and pleasure, grief and joy, passions and sensations succeed each other, and never all exist at the same time. It cannot, therefore, be from any of these impressions, or from any other, that the idea of self is deriv'd; and consequently there is no such idea (*Treatise*, 251-252).

In this passage, Hume wants to put a defender of the ontology of the self on the back foot by putting forth two demands, each of which may seem difficult to satisfy. Hume's first demand is that a defender of the ontology of the self must establish the impression that gives rise to the idea of the self. Hume's second demand is that even if it is supposed that a defender of the ontology of the self can show the impression that gives rise to the idea of the self, there is an intractable problem yet to be faced. That is, since an impression does not endure over time, Hume claims that any attempt to establish an idea of an

enduring/persistent self in variable impression will not be successful. Thus, either way Hume thinks that a defender of the ontology of the self has no option but to reconsider her position (cf. Robison in Tweyman 1995: 699). But the question remains: why would a defender of the ontology of the self accept the terms of the debate set by Hume? This question reminds us of our discussion of Hume's *psychologism* in § 3.4.2. If my conclusion in that section with respect to Hume's *psychologism* is correct, then we can adopt a similar line of critique to show why Hume's arguments against a defender of the ontology of the self will ultimately fail to succeed. But before we take up any such discussion, we need to probe further into Hume's reasons for challenging rival views of the self.

3.5.2 *Where is It?*

In this case, Hume's most famous objection against the realist ontology of the self comes in the following passage (also see chapter one § 1.1.2.2 A):

> For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* any time without a perception, and never can observe any thing but the perception…I am certain there is no such principle in me (*Treatise*, 252).

Here Hume puts *observability* as a precondition for the existence of the self. So he wants to test it (i.e, the self's observability) through engaging in an introspective investigation. One thing that is distinctive about Hume's introspective investigation of the self has to do with the fact that it is rooted in Hume's own first-person perspective. Yet despite Hume's first personal introspective investigation, the self did not turn up in perception. Hence, Hume concluded that there is no such thing we call a self. If Hume is right, then he has raised a powerful objection against a defender of the realist ontology of the self. Hume's objection targets a Cartesian self (Kim 2005: 31-35; see also Flew 1986: 94-98). However, we need to keep in mind that the objection also has a wider application that challenges any view that conceives a self as a distinct property bearer of some kind. Here I say, 'some kind,' in order to leave room for non-Cartesian ways of characterizing the notion of the self. In chapter six, I will discuss one such characterization of the self/person.

Most contemporary philosophers argue that Hume's failure to detect a self in experience undermines any positive support one can give for the realist ontology of the self. For example, Searle remarks, "Hume is absolutely right; there is no experience of this entity…," (2004: 203). Kripke also claims that "Descartes' notion seems to have been rendered dubious ever since Hume's critique of the notion of a Cartesian self," (Kripke 1980: 155 footnote # 77; see also Grossman 1965: 41-42; Ayer 1963: 83; Dennett 1991: Ch. 13; Armstrong 1980: 64; Westerhoff 2011: 66; cf. Blackburn 1999: 138-140). Despite an overwhelming company Hume enjoys from the contemporary philosophers, not everyone thinks that Hume's search for the self through introspection is even the right sort of approach to begin with.

For example, Shoemaker claims that Hume's denial may have some intuitive appeal to it. Yet for Shoemaker, the basis for Hume's denial is far from clear in terms of what exactly it means, or what its philosophical implications are. Assuming that the basis for Hume's denial is empirical, when Hume looks within for a self, what he ended up finding was instead only particular perceptions. In light of this, Shoemaker claims that Hume's denial that he is aware of a self can scarcely have the same basis as one's well-founded denial that one sees a tea kettle. For Shoemaker, the latter denial is well-founded only assuming that one has some idea of what it would be like to see a tea kettle. By contrast, Shoemaker claims that Hume is emphatic on the point that he has no idea of self *qua* subject of experiences nor does he think that he has any idea of what it would be like to introspect one. Shoemaker claims that despite the initial appearances, the basis for Hume's denial can hardly be empirical (Shoemaker in Cassam 1994: 119). In light of this, Shoemaker concludes, "If the basis of the Humean denial is less than clear, so also is its meaning. Sometimes it is put by saying that we are not 'acquainted' with a self or that we are not, in introspection, presented with a self 'as an object'. But what is it to be acquainted with something in the required sense?," (Ibid.). Chisholm also remarks:

> Our idea of 'a mind' (if by 'a mind' we mean, as Hume usually does, a person, or a self) is not an idea only of 'particular perceptions'. It is not the idea of the perception of love or hate and the perception of cold or warmth, much less an idea of love or hate and the heat or cold. It is an idea of that which loves or hates, and of that which feels cold or warm….That is to say, it is an idea of an *x* such that *x* loves or *x* hates

and such that *x* feels cold and *x* feels warm, and so forth (Chisholm 1969: 9; see also Russell 1912: 27).

Shoemaker and Chisholm in their own way touched upon two important points, which I already discussed in chapter one and also continue to discuss throughout this thesis. The first point is that, trying to look for a self as if it is an object is not only mistaken, it does not even make sense (see again chapter § 1.1.2.2 A & B). This is also Shoemaker's main criticism of Hume's introspective exercise in looking for a self. The other point has to do with the implication of the first person account of one's experience. Notice that as I pointed out earlier, Hume's report about the absence of a self in introspection, is rooted in Hume's own first person account. If so, as Chisholm rightly observed, Hume's denial of a self cannot be established without presupposing it. In chapter seven, I will be developing this idea in my own way to defend a realist ontology of the self.

On the other hand, in the *Critique of Pure Reason*, Kant agreed with Hume that the self might not be located in experience, given the changing nature of appearances. As Kant remarks, "consciousness of self…is merely empirical, and always changing. No fixed and abiding self can present itself in this flux of inner appearances," (A 107). However, unlike Hume, Kant insisted that one's experiences have to be unified. But such unity of experiences can only happen if all of one's representations are attributable to a single subject in order for them to qualify as one's own. Kant claims that to ensure the unity of experiences, there must be a condition which precedes all experiences that makes experience itself possible. Kant calls this a 'pure original unchangeable apperception' (*Ibid*.).[83] Furthermore, in the *Transcendental Deduction* B, Kant further argues why all of our representations must belong to a single subject. As Gardener sums it up:

> It is not that each of my representations must be actually accompanied by the reflection that it is mine, nor that I must be able to form a single thought comprehending all of my representations in one grand totality: Kant's claim is just that each of my representations must be such that it is possible for me to recognize it as mine in an act of reflection. For satisfaction of this condition, the representation

---

[83] Kant borrowed the term 'apperception' from Leibniz, for whom it meant perception of one's states (see Gardner 1999: 145).

'I', as invariant, a priori representation free from empirical content is essential; 'otherwise I should have as many-coloured and diverse a self as I have representations of which I am conscious'….'I think'—the necessity of the possibility of its accompanying my representations expresses the condition that any domain of objects must be conceived in the perspective of a thinking subject," (Gardner 1999: 146; see also Strawson 1966).

At this point, the question that comes to mind is this: Is Hume an eliminativist or a reductionist about the self? Based on the passage we saw at the beginning of this section, it is hard to categorize Hume in either domain. Yet we can get a clear answer for this question by looking at Hume's alternative proposal which, as we shall see, paves the way for Hume's deconstructionism of the self.

3.5.3 *The Bundle View of the Self*

Hume claims that once we set aside the ontology of the self that some metaphysicians defend, we can affirm the rest of mankind that they are nothing but a bundle or collection of different perceptions. Hume tells us that these perceptions succeed each other with an inconceivable rapidity and are in constant flux and movement. Our eyes cannot turn in their sockets without changing our perceptions (*Treatise* 252).[84] This is the clearest instance of Hume's reductionism of the selves/persons to bundles of perceptions. As we shall see in chapter five, Hume's reductionism of selves will also figure in the works of influential personal identity theorists like Parfit. Hume considers the mind as something that is being constituted by successive perceptions, which has no simplicity nor identity that we naturally attribute to it (*Treatise*, 253).

What do we make of Hume's claim that the self is a bundle of perception? Notice that in § 3.5.2, Hume told us that he could not detect a self through introspection. What this means is that Hume could not find the impression that matches with the impression(s) that is supposed to be constitutive of the self. The question then is: on what basis is Hume telling

---

[84] Here Hume's remarks are Heraclitean in their tone. A pre-Socratic philosopher, Heraclitus is credited with saying that everything is in a constant flux (Raven and Schofield 1957/1983: 195). However, unlike Hume, Heraclitus still maintains that, "men should try to comprehend the underlying coherence of things: it is expressed in the Logos, the formula or element of arrangement common to all things," (*Ibid*. 186).

us now that the self is a bundle of perceptions? What criterion (if any) has he used to distinguish self-constituting perceptions, from those that are not?[85] To my knowledge, I have not come across Hume's own answer for this question. Since nothing gives an underlying unity for the bundles that are supposed to constitute the self, it remains hard to make sense of how discrete perceptions come together to constitute a concrete object like a self. Moreover, it remains unclear what the nature of the relation that is supposed to exist between discrete perceptions. Hume's view in this regard is a precursor to contemporary bundle theory. As Loux explains:

> On this view [i.e., contemporary bundle view] familiar objects are complexes or wholes whose constituents are exhausted by those attributes that can be the objects of perceptual or introspective awareness. Denying the need for underlying subject for attributes, these empiricists have frequently invoked metaphors to express their analysis of the structure of concrete particulars. A concrete particular, we are hold, is nothing more than a "bundle," a "cluster," a "collection," or a "congeries" of the empirically manifest attributes that common sense associates with it (1998: 103).

But the question remains: "What is the ontological "glue" that holds the different items in each of these bundles together?" (Ibid). As we can see from the hitherto discussion, Hume's deconstructionism of a realist conception of substance and the self, has not in any clear way yielded the result Hume had expected. The question is: Will Hume's deconstructionism yield the result he wants when it comes to personal identity?

3.5.4 *Personal Identity*

Hume begins his discussion of personal identity by asking the following question: "What then gives us so great a propension to ascribe an identity to these successive perceptions, and to suppose ourselves possest of an invariable and uninterrupted existence thro' the whole course of our lives?", (*Treatise,* 253). Hume tells us that in order to answer this question, the issue of personal identity has to be approached from two aspects: (1) personal identity as it regards our thought or imagination; and (2) as it regards our passions or the

---

[85] This is an important question that often seems to have been passed by Hume's scholars, see e.g., Capaldi in Tweyman III (1995: Ch. 87); Penelhum (2003: Ch. 5); McIntyre in Tweyman III (1995: Ch. 93); Traiger in Tweyman III (1995: 92). However, in this regard, Pike attempts to defend Hume (1967: 159-165).

concern we have in ourselves (*Ibid*.). Hume discusses (1) in book I of his *Treatise*. He discusses (2) in book II of his *Treatise*. As already indicated at the beginning of this chapter, this thesis focuses on (1).

For Hume, the above question serves as a background assumption, in light of which he tries to show why we are mistaken in attributing identity both to ourselves as well as to other objects (*Treatise*, 253). As we recall, for Hume, identity is one of the philosophical relations. That is, as Hume explains, identity is the most universal of all relations, in that it is common to every being whose existence has any duration (*Treatise*, 14). However, Hume claims that when it comes to applying this relation to ourselves as well as to other things, we often make serious mistakes. What precisely are these mistakes and how is it that we end up making them? Before we look at this question, I want to point out two notions that are central to Hume's diagnoses of our mistaken conception of identity.

Hume claims that we have a distinct idea of an object which persists as invariable and uninterrupted through time. Hume calls this *identity* or *sameness*. Hume contrasts this notion of identity with the notion of *diversity*, which pertains to several different objects that exist in succession and are connected together by a close relation (*Treatise*, 253). Hume thinks that our fundamental mistake in thinking that objects endure through time arises from conflating the notion of identity and the notion of diversity with each other. But Hume's own claim that objects remain the same (i.e., without experiencing any change) through time is wrong. I already explained in chapter one that qualitative change, say, of the apple changing its colour from green to red (while remaining the same apple), can be a counterexample to the way Hume characterizes the notion of identity. Contra Hume, variation of time presupposes some sort of change, whether that is *qualitative*, *substantial* or *change of composition* (see again chapter one § 1.2.2 B). If so, every kind of change that objects experience through time, does not necessarily pose a threat to their identity (cf. Penelhum in Tweyman III 1995: 658).[86] In other words, identity is not incompatible with

_____

[86] Of course, at this point, it may be said that at the microscopic level, when elementary particles enter into an entangled state, there is no fact of the matter as to their identity. Lowe, for example, argues that this could be an instance of an ontic indeterminacy as opposed to epistemic or semantic (see e.g., Lowe 1994:

change (cf. Noonan 1999: 205). Having said that, however, Hume's own main claim is that our fundamental mistake in believing in the persistence of objects through time, emanates from attributing identity to objects (i.e., perceptions)—that exist only in succession and connected by a close relation. Hume's point is that we make a mistake when we confuse identity with relation. But what is the source of such confusion?

In this case, Hume blames imagination. For Hume, imagination is one of the faculties by which we repeat our impressions (*Treatise* 8). Hume claims that imagination makes us consider things to be uninterrupted and invariable. Thus, we mistakenly substitute the relation we witness among succeeding objects, for things that endure through time. As Hume remarks:

> The relation facilitates the transition of the mind from one object to another, and renders its passage as smooth as if it is contemplated one continu'd object. This resemblance is the cause of the confusion and mistake, and makes us substitute the notion of identity, instead of that of related objects. However at one instant we may consider the related succession as variable or interrupted, we are sure that the next to ascribe to it perfect identity, and regard it as invariable and uninterrupted. Our propensity to this mistake is so great from the resemblance above-mention'd, that we fall into it before we are aware….In order to justify to ourselves this absurdity, we often feign some new and unintelligible principle, that connects the objects together, and prevents their interruption or variation. Thus we feign the continu'd existence of the perceptions of our senses, to remove the interruption; and run into the notion of a *soul*, and *self*, and *substance*, to disguise the variation (*Treatise*, 254).

This passage once again sums up Hume's deconstructionism in the context of his discussion of personal identity. In this passage, we also see that Hume's *psychologism* comes in full force. Hume's focus is on diagnosing how we mistakenly attribute identity to objects of our perception on the one hand and to ourselves on the other hand. Notice that Hume is consistently applying what I introduced in § 3.4, as the three-prong objections, namely *the Distinction Objection*, *the Knowledge Objection* and *the Origin Objection.* As we recall, the first objection deals with the distinction between the self and its properties. The second

---

110-114). Here my point only applies to ordinary familiar objects at the macroscopic level (e.g., selves, trees, etc.). So my point still stands.

objection deals with our knowledge of the impressions that gives rise to the self; and the third objection deals with our lack of specifying which impressions are responsible for the idea of substance and the self. Regarding identity, Hume remarks:

> For when we attribute identity, in an improper sense, to variable or interrupted objects, our mistake is not confin'd to the expression, but it commonly attended with a fiction, either of something invariable and uninterrupted, or of something mysterious and inexplicable, or at least with a propensity to such fictions (*Treatise*, 255).

In talking about the identity we ascribe to persons, Hume remarks, "the identity, which we ascribe to the mind of man, is only a fictitious one," (*Treatise*, 259). This means that, as Hume sees it, the identity we attribute to persons is not different from the identity we attribute to plants, animals, ships and houses. In each of these cases, there are changes that take place. For example, trees lose the matter that composes them. Houses get demolished and could be rebuilt out of completely different materials. Despite such changes, we often freely attribute identity to these objects, even when none of the objects are the same with the original objects (see e.g., *Treatise*, 256-258). However, here Hume seems to be conflating two notions of substance. As we recall, in chapter one, I distinguished between the *stuff/matter* notion of substance on the one hand and the notion of substance taken as a fundamental existent or *ousia*, on the other hand.

Taken in the first sense of substance, for example, a tree can lose the particles that compose it at any given time without losing its identity. That means that the stuff/matter that composes a tree should not be equated with the tree itself. I can say that the tree in my backyard is the same tree that I have known since I came to England. But since I came to England, the tree in my backyard has recycled several times the stuff that composes it—as a matter of biological process. Yet, I am justified in saying that the tree in my backyard is the same tree that endures the loss of stuff that composes it from time to time. Similarly, persons endure through mereological replacement. As Moreland remarks, "we are enduring continuants even though we undergo various changes and our bodies experience part replacement," (2009: 115). Of course, mereological replacement in the case of material objects (e.g., ship, houses) is extremely controversial, which does not concern us for present purposes (see e.g., Lowe 2002: Ch. 2). But Hume's denial of strict identity to living creatures

on the basis of the loss of material composition is clearly a non-starter. In this regard, as we saw in chapter two, Locke is much more careful.

Yet against all such considerations, Hume remarks:

> The identity, which we ascribe to the mind of man is only a fictitious one, and of a like kind with that which we ascribe to vegetables and animal bodies. It cannot, therefore, have a different origin, but must proceed from a like operation of the imagination upon like objects (*Treatise*, 259).

Hume claims that there is no difference between the identity of persons and the identity of other things, such as artefacts, plants and animals. In this case, Hume's claim is extremely influential in the contemporary personal identity debate. For example, as I explained in chapter one, the complex view theorists follow Hume in this regard (see e.g., Shoemaker in Gasser and Stephen 2012: Ch. 6). Hume further strengthens his point as follows:

> 'Tis evident, that the identity, which we attribute to the human mind, however perfect we may imagine it to be, is not able to run the several different perceptions into one, and make them lose their characters of distinction and difference, which are essential to them. (*Treatise*, 259).

Hume's point is that distinct and different perceptions that exist are not united by any underlying principle of any kind (cf. *Treatise*, 260). Hume thinks that there is no way for the human mind to observe "any real connexion among objects, and that even the union of cause and effect, when strictly examin'd, resolves itself into a customary association of ideas," (*Treatise* 259-260). Hence, Hume concludes,

> 'Tis therefore, on some of these three relations of resemblance, contiguity and causation, that identity depends; and as the very essence of these relations consists in their producing an easy transition of ideas; it follows, that our notions of personal identity, proceed entirely from the smooth and uninterrupted progress of the thought along a train of connected ideas, according to the principles above-explain'd. (*Treatise*, 260).

For Hume, an uninterrupted progress of our thought, which we attribute to the successive existence of a mind or a thinking person, is produced by resemblance and causation. Unlike Locke who thought that memory plays a critical role in personal identity, Hume briefly argued that memory only discovers personal identity (*Treatise*, 262). That means that memory plays only an evidential role in a sense of allowing us to have knowledge of our

own identity. In light of the hitherto considerations, Hume claims that questions about personal identity are grammatical rather than philosophical difficulties. Hume also says that all the disputes besetting the identity of connected objects are merely verbal in so far as they lead us to some fiction or imaginary principle of union (*Treatise*, 262). Here Hume's remarks represent the climax of his deconstructionism, where he announces that no substantive discussion can be had with respect to the principle of union (i.e., substance).

3.5.5 *Hume's Confessions*

Hume applied his deconstructionism to *substance*, the *self*, and *identity/personal identity*. As we saw, Hume had nothing positive to say about these things (Cf. Wolfram 1974: 586-593). Where does all that leave Hume's deconstructionism? At this point, it is important for us to briefly look at the implications of Hume's own confessions in the *Appendix* (see *Treatise*, 633-636). As I have already stated at the beginning of this chapter, I am not interested to engage in contemporary speculation regarding the reasons for Hume's confession in the *Appendix*. These speculations are still ongoing. My own take on Hume's confession in the *Appendix* is different from orthodox approaches. I take Hume's words in the *Appendix* at face value. That is, Hume was deeply dissatisfied with his theory of personal identity.

The first thing to notice in the *Appendix* is that Hume hasn't abandoned his position on substance, the realist ontology the self, and identity/personal identity (see e.g., *Treatise* 633-635). In my view, what is distinctive in the *Appendix* about Hume's confession has to do with his realization that his deconstruction project comes at a price. For example, as Hume describes:

> But upon a more strict review of the section concerning personal identity, I find myself involv'd in such a labrynth, that, I must confess, I neither know how to correct any former opinions, nor how to render them consistent (*Treatise*, 633).

In this particular confession, Hume is telling us that he doesn't know how to correct his previous opinions concerning personal identity. In this regard, for example, Hume rejected that identity is not a fact that holds in the nature of things. Rather, identity is something we

conjure up in our imagination. Hume himself understood that deconstructing the *uniting principle* (i.e., substance) was not as simple as he had first thought. As Hume remarks:

> In short there are two principles, which I cannot render consistent: nor is it in my power to renounce either of them, viz *that all our distinct perceptions are distinct existences,* and that the mind never perceives any real connexion among distinct existences. Did our perceptions either inhere in something simple and individual, or did the mind perceive some real connexion among them, there wou'd be no difficulty in the case (*Treatise*, 636).

Again what we see in this confession is that Hume's deconstructionism has no advantage by way of helping us make sense of our experience of the world around us in general, and in our own case, as subjects of experience, in particular. Ironically, in light of Hume's confession, it seems fair to say that to have a defensible ontology of the self on the one hand, and the metaphysics of identity/personal identity on the other, we still need substance ontology.

3.6 Summary

In this chapter, our main focus has been on Hume's discussion of substance, the self and identity/personal identity. In this regard, we saw that Hume's discussion is negative. We also saw in detail what I called Hume's deconstruction project. Hume applied his deconstructionism against three main traditions, namely, *the Aristotelian*, *the Lockean* and *the Cartesian*. We also saw that Hume's epistemology is rooted in his strong empiricism. In light of this, Hume rejected traditional metaphysical claims about the ontology of the self. We also saw that Hume's rejection of traditional metaphysics in favor of his own strong empiricism has not proven to be entirely helpful when it comes to issues such as substance ontology, and identity/personal identity. As Hume's own confession in the *Appendix* seems to indicate, jettisoning metaphysics may not be as simple as it initially might seem to be. If this is right, then such a finding all the more strengthens the background assumption of this thesis, which is the importance of traditional metaphysics for the ontology of the self. That said, in part three of this thesis, my discussion will focus on the contemporary issues that deal with the methodology of thought experiments and the personal identity debate. In chapter four, I will take up the first issue and the second issue will be discussed in chapter five. In chapter six, I will discuss issues related to the self in philosophy of mind.

Part III: Contemporary Debates on Personal Identity

Chapter four

Thought Experiment and Personal Identity

In chapters two and three, my main focus has been in showing the historical relevance of Locke and Hume in relation to the contemporary personal identity debate on the one hand, and their account of substance ontology in relation to the self/person on the other hand. In this case, one of the things I tried to emphasize was that our interest in Locke and Hume is not confined just to their historical significance. Rather their influence has continued well into our time. Both of these philosophers, in their own way, have set the tone for the contemporary personal identity debate. As we shall see in chapter five, Locke's legacies can be seen in the neo-Lockean tradition; whereas Hume's legacies can fairly be attached to the contemporary reductionist account of personal identity. We will have a chance to see concrete examples of each of these instances in chapter five. That said, in this short chapter, my main aim is to tackle the issue of methodology in the personal identity debate with a particular focus on the neo-Lockean tradition.

Initially, how this chapter fits into the overall dialectic of the thesis may not be clear. But for reasons we shall see, a discussion of this issue is crucial, since it is related to the discussion I will be undertaking in chapter five on the contemporary debates on personal identity on the one hand, and the negligence of substance ontology on the other. For example, one of the ways most contemporary personal identity theorists try to justify their marginalization of substance ontology is by appealing to thought experiments. Of course, in this regard Locke himself opened the door by denying the centrality of substance ontology for personal identity via his soul-swap thought experiment. However, as I will attempt to show in this chapter, it would be problematic to use thought experiments as a methodological proper to argue for personal identity.

In light of this, my bringing the issue of methodology right before chapter five is a deliberate strategic move on my part. As I already indicated in chapter one (see § 1.2 footnote # 27), I want to suggest my own way with respect to how to best utilize puzzle cases in the context of personal identity debates. In this case, I will emphasize three things.

A. The role thought experiments must be allowed to play.

B. The constraints within which thought experiments must be exercised.

C. The place thought experiments deserve and do not deserve.

I shall call (A) the Role Condition; (B) the Constraint Condition; and (C) The Priority Condition, respectively. In § 4.5, I will explain what (A)-(C) amount to. I will also argue why it is important for thought experiments invoked in personal identity discussions to satisfy (A)-(C). In light of this, whatever I say in this chapter concerning the use of puzzle cases will be confined to personal identity. So the remarks I make may not be applicable to the use of thought experiments in other contexts, say for example, in scientific contexts or otherwise. Moreover, the discussion I advance in this chapter will not be exhaustive. As indicated earlier, my interest in this short chapter is not to discuss the methodology of thought experiments for its own sake.[87] Rather, my goal is specific and aimed at paving the way for the discussion I will advance in chapter five which concerns with what is being prioritized in the contemporary personal identity debates on the one hand, and what is being pushed aside on the other.

In this case, as we recall, in chapter one (see § 1.2 and 1.2.2 D), I had expressed my suspicion that the lack of consensus in the contemporary personal identity debates, can in part, if not in whole, be attributed to the way thought experiments are often utilized in thinking about the problem of personal identity. But how are thought experiments used in personal identity debates? This is the question I will attempt to answer in what follows. In attempting to answer this question, my goal is to reiterate the point I have been emphasizing in this thesis, which is, in thinking about the problem of personal identity, our starting point must be what I called the *Aristotelian Question* that takes the form: '*what is X*?' As we shall see in chapter five, it remains very difficult to see what role (if any) the *Aristotelian Question* plays, when the neo-Lockeans discuss about the problem of personal

─────────────────────────

[87] For an excellent discussion on thought experiments quite in general (see Sorensen 1992; Horowitz, and Massey, eds. 1991; Brown and Fehige 2008). For thought experiments in natural science (see Brown 2011). For thought experiments in philosophy (see Rescher 2005; Haggvist 1996).

identity. In this case, the problems I had located in Locke's own account of the ontology of persons (see again chapter two § 2.2 C), equally afflict the modern-day neo-Lockeans' account of the ontology of persons. I will take up this discussion in chapter five.

4.1 *Thought Experiments in the Neo-Lockean Tradition*

What is distinctive about the neo-Lockean tradition is that, not only is it currently the most popular approach in personal identity debates, but it is also the most pro-thought experiment tradition (Rovane 1998: 35). Of course, the non neo-Lockeans also appeal to puzzle cases in debating the problem of personal identity. But when the non neo-Lockeans do that, as we shall see, their discussion is often framed in terms of a response given against the neo-Lockeans' view of the Psychological Criterion for personal identity (see e.g., Olson 1997: Ch. 3; Williams 1973: Chs. 2-4). So in this case, the neo-Lockeans are primarily responsible for the discussions often generated in the literature. No doubt that the neo-Lockeans drew their inspiration to appeal to puzzle cases from Locke's '*the prince* and *the cobbler*' thought experiment. As we saw in chapter two, this thought experiment focuses on the consequence of "soul swap" for the identity of persons. That is, if the soul of the prince enters the body of the cobbler and vice versa, then the prince will be identified with the cobbler's body and *ditto* for the cobbler. This means that given Locke's view, "a person goes wherever his/her soul goes". Of course, this does not mean that for Locke, a person is a soul; rather the soul is a bearer of a person (see again chapter two). Although the neo-Lockeans do not follow Locke in embracing the notion of the soul, as we shall see, they make a similar Lockean point by substituting the "soul swap" scenario for a "brain swap" scenario.

In this case, the neo-Lockeans' focus on well-known cases of brain transplant, scattered existence (i.e., brain stored in a vat), bionic replacement, teletransportation, fission and the like (see e.g., Garrett 1998: 16-17; Shoemaker 1984: Part Two; Parfit 1984: Part Three; also see Shoemaker 1963: 22-25). That said, however, unlike his modern day followers, Locke himself did not give a prominent place for thought experiments. His brief discussion of the prince and the cobbler thought experiment was not meant to set a methodology proper for how to think about the problem of personal identity. As we saw in chapter two in § 2.2 A,

Locke himself opened his discussion of personal identity with the *Aristotelian Question*. But as we shall see, the neo-Lockeans do not seem to have learned this crucial lesson from Locke.

4.2 *What are Thought Experiments?*

Philosophers describe the key features of thought experiments in philosophy by using various terms and expressions. For example, Williamson calls them "the armchair methods," (2007: 179). Brown and Fehige call them, "devices of the imagination," (1996/2014). Robinson describes them as, "tools" employed in conceptual analysis (2004: 537). Sorensen compares them to "compasses" used for determining direction (1992: 288). However, as they stand, these expressions are not informative enough in telling us about how the methodology of thought experiments work. In this case, in her famous book entitled *Real People: Personal Identity without Thought Experiments*, Wilkes explains the central idea behind the thought-experimental method as follows:

> Suppose that we want to test a claim made by some scientific theory (and hence to test the theory); suppose we want to see what might follow if certain theoretical claims were true; suppose we want to examine the range and scope of a concept. It may be appropriate, in all these different domains, to ask a 'what if...?' question. Such a question typically postulates an imaginary state of affairs, something that does not in fact happen in the real world. ...By such means...we may, perhaps, get a weaker or stronger reason for thinking a scientific claim to be true or false; for claiming a discovery about the limitations and scope of one of our everyday concepts (1988: 2).

Although Wilkes's characterization of the methodology of thought experiments focuses on a scientific theory, her remarks are equally applicable to thought experiments in philosophy. Philosophers also appeal to imaginary cases to see what might follow if certain imaginary state of affairs were true. They ask questions such as "Would it be the same *F* if *p*?" where *p* states some sort of science fiction scenario (Robinson 2004: 537). Philosophers attempt to answer such questions on the basis of the sort of responses evoked by imaginary cases, in light of which they also attempt to show the limitations and scope of our everyday concepts/beliefs concerning a particular issue. For example, as Gendler remarks:

> By presenting content in a suitably concrete or abstract way, thought experiments... evoke responses that run counter to those evoked by alternative presentations of

relevantly similar content. But exactly because of this, the responses they evoke may well remain in disequilibrium with responses evoked in alternative ways. When thought experiments succeed as devices of persuasion, it is because the evoked response becomes dominant, so that the subject comes (either reflectively or unreflectively) to represent relevant non-thought experimental content in light of the thought experimental conclusion (2007: 86).

As Gendler sees it, thought experiments evoke responses that would eventually lead one to embrace a certain conclusion regarding a particular issue. In this case, most personal identity theorists take Gendler's remarks on board, since they too believe that imaginary cases are used to elicit responses. For example, in this case, one of the most popular and influential imaginary cases in personal identity literature is Parfit's Teletransporter. Parfit opens his discussion on personal identity with the Teletransporter thought experiment (1984: Ch. 10). Parfit's imaginary story has two parts. In the first part, Parfit's remarks:

> I enter the Teletransporter. I have been to Mars before, but only by the old method, a space-ship journey taking several weeks. This machine will send me at the speed of light. I merely have to press the green button. Like others, I am nervous. Will it work? I remind myself what I have been told to expect. When I press the button, I shall lose consciousness, and then wake up at what seems a moment later. In fact I shall have been unconsciousness for about an hour. The Scanner here on Earth will destroy my brain and body, while recording the exact states of all of my cells. It will then transmit this information by radio. Travelling at the speed of light, the message will take three minutes to reach the Replicator on Mars. This will then create, out of new matter, a brain and body exactly like mine…Examining my new body, I find no change at all. Even the cut on my upper lip, from this morning's shave, is still there (1984: 199).

In the second part, Parfit remarks:

> Several years pass, during which I am often teletransported. I am now back in the cubicle, ready for another trip to Mars. But this time, when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I leave the cubicle, and say to the attendant: 'It's not working. What did I do wrong?' 'It is not working,' he replies, handing me a printed copy. This reads: 'The New Scanner records your blueprint without destroying your brain and body….But it seems to be damaging the cardiac systems which it scans…though you will be quite healthy on Mars, here on Earth you must expect cardiac failure within the next few days…,' (*Ibid*. 199-200).

What do we make of Parfit's imaginary story? Given the fanciful nature of the thought experiment, one could be tempted to dismiss it outright. But Parfit disagrees. This is

because since the imaginary story arouse in most of us strong beliefs, Parfit claims that one cannot be justified in dismissing it. For Parfit, these beliefs are not about our words; rather, they are about ourselves. Parfit claims that imaginary cases allow us to discover what is involved in our continued existence over time as the same people. That is, imaginary cases expose what we believe about the nature of personal identity over time (1984: 200). In light of this, in the first part of the imaginary story above, Parfit's point is that even though the Scanner destroys one's brain and body, out of the blueprint that is beamed to Mars, another machine makes one's organic *Replica* (*Ibid*.). In the second part of the imaginary story, even though the Scanner does not destroy one's body and brain, through the blueprint sent to Mars one's Replica is produced. But after a little while, one will suffer cardiac arrest. But in such case, one's Replica will take up one's life where one leaves off (*Ibid*.). So after one's death here on Earth, one's life and that of one's Replica overlap. Parfit calls this the "*Branch-Line Case*," (*Ibid*. 201).

Parfit's imaginary story elicits different responses. For example, (i) some might take it to be the fastest way of travelling to Mars (*Ibid*. 200); (ii) another person might say that no human being could survive teletransportation (see e.g., Johnston 1987: 64); and (iii) Parfit's own end goal is to convince his readers that, insofar as one takes having one's Replica as good as ordinary survival, it remains entirely irrelevant to insist that "numerical identity is what matters in survival" (see e.g., *Ibid*. Chs. 11-12). But how should one adjudicate over the responses in (i)-(iii)? Notice that, of (i)-(iii), the most serious response is the one that is given in (iii). This is because, (iii) is not a simple response; rather it is a serious metaphysical claim about the state of one's identity. But why should we think that the Teletransporter story is adequate enough to convince us to embrace (iii)? Why can't one settle for (ii) instead? For reasons we shall see in § 4.2-4.4, establishing the responses elicited on the basis of thought experiments in philosophy will not be as straightforward as most philosophers often tend to think. As we shall see, in this regard, not everyone receives the conclusions that are based on imaginary stories with open arms.

4.3 *Contemporary Philosophers' Attitude toward Thought Experiments*

As Nozick once remarked, "so many puzzling examples have been put forth in recent discussions of personal identity that it is difficult to formulate, much less defend, any consistent view of identity and non-identity," (1981: 29). In their essay entitled, "When are Thought Experiments Poor Ones?" Peijneburg and Atkinson also remark:

> One of the things that sets contemporary analytic philosophy apart from its older variant is its ample use of thought experiments. While early analytic philosophers like Russell, Ayer or Carnap seldom rely on this kind of hypothetical reasoning, modern ones like Jackson, Searle and Putnam do not eschew the most bizarre accounts of Zombies, swapped brains, exact *Doppelgänger*, and famous violinists who are plugged into another body (2003: 305).

In light of Nozick-Peijneburg-Atkinson's observations, one may wonder why contemporary philosophers put much emphasis on thought experiments.[88] Whatever other reasons may also be, it is undeniable that the use of puzzle cases seems to be contemporary philosophers' most preferred methodology. This is no less evident in personal identity literature. As I briefly pointed out in § 4.1, in the context of the contemporary personal identity debate, the challenge one often faces is that of showing what justifies one's judgment of a particular puzzle case as being more truth revealer about personal identity compared to the other. This problem has not been explicitly acknowledged by most contemporary personal identity theorists. But those philosophers who understood the problem, suggest what seems for some to be a radical position (see e.g., Coleman 2000). In this case, for example, Wilkes argues why we should discuss personal identity without thought experiment (1988). This is because as Wilkes argues, thought experiments in philosophy suffer from what she calls the lack of the 'background conditions.' In explaining the background conditions, Wilkes remarks:

> The experimenter—any experimenter, in thought or in actuality—needs to give us the background conditions against which he sets his experiment. If he does not, the results of his experiment will be inconclusive. The reasons for this is simple and obvious: experiments, typically, set out to show what difference some factor makes; in order to test this, other relevant conditions must be held constant, and the

---

[88] In the context of personal identity discussions, the inspiration for emphasizing on thought experiment comes from Locke's the cobbler and the prince soul-swap scenario. In this regard, most philosophers find the methodology of thought experiment liberating in a sense of giving them unrestricted ways to entertain various sorts of possibilities with respect to the ground of personal identity.

problematic factor juggled against that constant background. If several factors were fluctuating, then we would not know which of them (or which combinations of them) to hold responsible for the outcome (1988: 7; cf. Popper 1959: Appendix XI).

Here Wilkes puts the bar very high for those who use thought experiments in philosophy. Philosophers cannot meet the background conditions in constructing their thought experiments in anything like the way Wilkes requires. This is because, as we saw earlier, philosophical thought experiments are "armchair methods," (Williamson 2007: 179). However, despite being armchair methods, as Williamson further remarks, "much of the philosophical community allows that a judicial act of imagination can refute a previously well-supported theory (*Ibid*.). At this point, it is not hard to see that there is a difference between scientific thought experiments and philosophical thought experiments (see e.g., Peijneburg and Atkinson 2003). In this case, Wilkes's remarks are more fitting in the context of scientific thought experiments. This is because, her remarks are more applicable to experiments done in the lab.

At this point, some philosophers may claim that they should be left alone to use any thought experiment in any way they like, since they are not doing scientific experiment (cf. Schlesinger 1993: 97). But Wilkes disagrees. Whether or not they are scientific or philosophical, Wilkes argues that all thought experiments must satisfy the background conditions. That means that philosophical thought experiments cannot be let off the hook. Wilkes defends this point by using a popular thought experiment of 'persons splitting like amoebae' or what is known as a fission thought experiment. Wilkes argues that for this thought experiment to work, the relevant background information must be specified. As Wilkes remarks:

> It is obviously and essentially relevant *to the purpose of this thought experiment* to know such things as: how often? Is it predictable? Or sometimes predictable and sometimes not, like dying? Can it be induced, or prevented? Just as obviously, the background society, against which we set the phenomenon, is now mysterious. Does it have such institutions as marriage? How would that work? Or universities? It would be difficult, to say the least, if universities doubled the size every few days, or weeks, or years. Are pregnant women debarred from splitting? The *entire* background here is incomprehensible. When we ask what *we* would say if this happened, who, now, are 'we'? (*Ibid*. 11).

To my knowledge, no contemporary personal identity theorist provided the sort of background information Wilkes's requires for the fission thought experiment (see e.g., Parfit 1971: 3-27; Merricks 1997: 163-186; Garrett 1998: Ch. 4; Martin 1998: 13-15, Chs. 2-3). Philosophers who take the fission thought experiment seriously claim that it reveals our deep-seated beliefs about what matters in survival. For example, Garrett remarks, "it shows, that the identity of a person can be extrinsically grounded. That is, the identity of a person may be fixed by the existence of another, causally unrelated, person," (1998: 58). Here Garrett seems to be echoing Nozick's, 'the closest continuer theory.' Nozick claims that the closet continuer view presents a necessary condition for identity (1981: 34). As Nozick explains, "something at t_2 is not the same entity as x at t_1 if it is not x's closest continuer. And "closest" means closer than all others; if two things at t_2 tie in closeness to x at t_1, then neither is the same entity as x," (*Ibid*.). In this regard, in chapter five, I will briefly discuss Parfit's and Lewis's exchange on the implications of a fission thought experiment for numerical identity over time (see e.g., Parfit 1971; Parfit in Rorty 1976: IV; Lewis in Rorty 1976: I). But if one sides with Wilkes and takes her requirement for background condition seriously, then it would be very difficult to see the basis for endorsing a highly counterintuitive proposal about a person's identity being extrinsically grounded. If the very thought experiment itself is severely under-described, what compelling reason will there be for one to embrace it?

However, Garrett argues that thought experiments help us to solve the problem of ordinary identity judgments. For example, Garrett claims that when we judge that the speaker before us now is identical to the speaker who began speaking two hours ago, we typically make this judgment of identity under optimal conditions. This is because, as Garrett points out, we can observe that the earlier person is both physically as well as psychologically continuous with the later person. In this everyday case, our basis for our judgment of identity is the obtaining of both physical and psychological continuities. However, Garrett says that our ordinary judgments of identity will not help us to determine which continuity (i.e., the body or the brain) is central to the identity of a person over time (1998: 14). To determine this matter, Garrett suggests that we must consider thought experiments, where these continuities come apart. Although the thought experiments we use may be

technically impossible (even may always be so), Garrett claims that there is no reason to think that most popular thought experiments in personal identity literature are physically impossible, in a sense that they are inconsistent with the laws of nature. In this case, Garrett discusses thought experiments such as brain transplant, teletransportation, scattered existence (i.e., brain in a vat), bionic replacement, fission, indeterminacy and others (1998: 15-17). Garret claims that none of the above thought experiments is logically impossible (*Ibid*. 17).

Of course, it could well be argued that thought experiments are suggestive of various sorts of possibilities with respect to some states of affairs. In this case, four types of possibilities are commonly discussed in the literature. For example, in their *Conceivability and Possibility*, Gendler and Hawthorne explain these possibilities as follows: The first possibility is epistemic that takes the form, ''for all I know, thus-and-such obtains.'' The second possibility is nomological. That is to say that if *P* is said to be nomologically possible, then *P* must be consistent with the body of truths expressed by those laws—whether those be the laws of physics, biology or otherwise. The third possibility is logical. For example, *P* is said to be logically possible, if *P* does not violate the standard rules of logical inference. The fourth possibility is metaphysical, which is taken to be the most fundamental conception of possibility. This is often expressed by using locutions such as 'how things might have been,' (2002: 3-5).

Of the four types of possibility above, the various puzzle cases used in personal identity debate are often linked to logical/metaphysical possibility. Once the notion of possibility is invoked, the notion of conceivability naturally follows. For example, Chalmers in his well-known essay entitled, "Does Conceivability Entail Possibility?" points out that it is a common practice in philosophy to use *a priori* methods to draw conclusions about what is possible and what is necessary, and in doing so to draw conclusions about matters of substantive metaphysics. Chalmers further points out that arguments like this proceed in three steps: (i) one begins with an epistemic claim regarding what can be known or conceived; (ii) from there to a modal claim regarding what is possible or necessary; and (iii) from there to a metaphysical claim regarding the nature of things in the world (in Gendler

and Hawthorne 2002: 145). In this regard, as Chalmers claims, the methodology of conceivability plays a central role (*Ibid*. 146).

Chalmers claims that if one argues that some states of affairs are conceivable, then one can conclude that the conceived states of affairs are possible. For Chalmers, the kind of possibility invoked here is metaphysical possibility, as opposed to physical possibility, natural possibility, or other sorts of possibility (*Ibid*.). In light of this, for Chalmers, the idea that conceivability can act as a guide to metaphysical possibility is not an implausible idea. However, Chalmers claims that to argue that conceivability entails physical or natural possibility is clearly implausible. For example, Chalmers claims that it seems conceivable that a particular object could travel faster than a billion meters per second. Yet this hypothesis is hardly physically and naturally possible, since it contradicts the laws of physics and the laws of nature. But if we suppose a metaphysically possible world with different laws, then such a hypothesis would be possible. This is a world, say for example, God might have created, if he had so chosen. If such a world had been created, then an object's travelling faster than a billion meters per second would have been possible. So in this case, Chalmers concludes that despite the fact that conceivability does not mirror natural possibility, it certainly mirrors metaphysical possibility (*Ibid*.). Chalmers's essay discusses the relation between conceivability and possibility at great length; but for present purposes the details do not matter (see further *Ibid*. 145-200).

Having said that, Chalmers's 'conceivability entails possibility' proposal will have its own drawbacks, if it is seen from the standpoint of the personal identity debate. But as we shall see in § 4.4, one of the drawbacks of appealing to conceivability as a guide to possibility has to do with the fact that it encourages the use of puzzle cases without constraint. Moreover, even if it could be said that, say for example, *P* is conceivable, it could also be argued that *P's* conceivability does not guarantee necessarily its actual possibility. Here following Bealer, we could ask: "Why should your conceiving that *P* provide you with evidence that *P* is possible?" (in Gendler and Hawthorne 2002: 76). As we shall see in chapter five, how one answers this question will have significant bearing on how one sees the role of various puzzle cases often invoked in personal identity debates.

Some philosophers like Schlesinger argue that, even if thought experiments may have some constraints placed on them in science, no such constraints are required in metaphysics (1983: 97). I disagree with Schlesinger's claim. As I will argue in § 4.4, my view is that thought experiments must be utilized under some form of reasonable constraint. At this point, it may be asked, what counts as a reasonable constraint. Again in this chapter, it is not my goal to suggest any criterion by way of answering this question. However, in light of the three conditions I mentioned at the beginning of this chapter, namely the Role Condition, the Constraint Condition and the Priority Condition, I will make my own suggestions concerning how thought experiments should be used. I will take up this task in § 4.5. But before I do that, in § 4.4 I will briefly discuss the notion of intuition.

4.4 *Intuition and Thought Experiments*

Garrett claims that one of the criticisms often raised against the use of thought experiments has to do with appealing to intuitions. As Garrett points out, it has been claimed that we should not take our intuitions about thought experiments as guides to philosophical truth, since such intuitions may be prejudiced and unreliable (1998: 14). Notice that Garrett's remarks presuppose that intuitions can be guides to philosophical truth. Before we examine Garrett's claim, it will be helpful to understand what intuitions are. Goldman claims that, when philosophers attempt to decide the concept of something, say of, identity, they often consider actual and hypothetical examples. In doing so, they ask whether these examples provide instances of the target category or concept (2007: 1). The mental responses people give to these examples can be called, "intuitions" (*Ibid.*). Goldman further notes that these intuitions can be used as evidence for the correct answer (*Ibid.*). Bealer also characterizes intuition as follows:

> Intuition is the source of all a priori knowledge—except, of course, for that which is merely stipulative. The use of intuitions as evidence (reasons) is ubiquitous in our standard justificatory practices in the a priori disciplines—Gettier intuitions, twin-earth intuitions, transitivity intuitions, etc. By intuitions here, we mean *seeming*s: for you to have an intuition that A is just for it to *seem* to you that A. Of course, this kind of seeming is *intellectual*, not experiential—sensory, introspective, imaginative. Typically, the contents of intellectual and experiential seeming cannot overlap. You can intuit that there could be infinitely many marbles, but such a thing cannot seem experientially (say, imaginatively) to be so. Intuition and imagination are in this way distinct (Bealer in Gendler and Hawthorne 2002: 73).

Although Bealer construes intuitions using the language of 'seemings,' for him such seemings are not phenomenological. That is, when someone intuits something, he/she does not have an experience of what it is like to intuit that particular thing. In this sense, intuitions are not experience based. Rather we come to have certain intuitions on the basis of grasping something purely on an intellectual basis. However, it is not entirely clear to me whether Bealer's characterization exhausts all that there is to intuitions. At least, in the context of the personal identity debate, philosophers often seem to take intuitions as judgments one makes concerning the possibility or impossibility of some states of affairs. In this case, my own understanding of the notion of intuition goes along with Goldman's characterization of "intuitions" above. More specifically, we can also take intuitions in the words of Van Inwagen as "our beliefs," (1990: 169). If we adopt Van Inwagen's suggestion, then contra Bealer, the beliefs we hold about something, in this case, personal identity cannot be divorced entirely from experience. Taken this way, intuitions can also be taken to be person-relative, that is to say that they are not universal. In my view, that is why different people react differently to thought experiments such as Parfit's Teletransporter.

If this is true, then the question remains. Can we settle a particular debate that is elicited on the basis of thought experiments by appealing to our intuitions? The answer for this question seems to be far from clear. Suppose that Parfit and I engage in a heated debate over the merit of the Teletransporter in settling the question of whether or not numerical identity is what matters in survival. As it stands, this question is still divisive among contemporary philosophers (see e.g., Sosa 1990: 297-322; Johnston in Dancy 1997: Ch. 8). Most (if not all) philosophers still struggle to accept Parfit's conclusion about whether or not identity matters in survival. Suppose that I am among those philosophers who struggle to embrace Parfit's conclusion. How are we going to settle the debate between Parfit and myself? My own view is that the disagreement between Parfit and myself would be irresolvable. The reason is that there does not seem to be any principled way to show whether my intuition is superior to that of Parfit's and vice versa.

In saying this, however, I am not implying that there is no fact of the matter as to whose intuition is right. There may be. But how can we tell? Since intuitions seem to be person-

relative, it seems hard to see how merely rooting them in puzzle cases will do the kind of job some personal identity theorists like Parfit hope they do. My point is that intuitions can go either way. That is to say that any given two persons' intuitions with respect to a certain thought experiment could be different. For example, Swinburne defends the simple view of personal identity by appealing to the brain hemisphere transplant puzzle case, in light of which he attempts to show why neither the continuity of a mental life nor a bodily constitution is necessary for personal identity (see e.g., Swinburne in Gasser and Stefan 2012: Ch. 5). On the other hand, Salmon defends a brain based view of personal identity by appealing to a brain swap thought experiment, in light of which he attempts to show why neither the body nor the psychology is necessary for personal identity (see e.g., Salmon 2006: Ch. 11; see also Van Inwagen 1990: 169; Williams 1973: Ch. 4). This is one reason why it is exceedingly difficult to resolve controversies in personal identity discussions. If this is correct, then Garrett's earlier claim that intuitions are guides to philosophical truth cannot be taken without qualification.

4.5 What to do then?

The more thought experiments are fanciful in their nature, the bigger the chances are for diverse intuitions to be elicited. So to narrow the gap between diverse intuitions, we need to come up with the sorts of thought experiments that are in some sense close to our experience of the world around us as well as other fellow humans. In such cases, the chances for thought experiments to elicit positive intuition will most likely increase. Suppose that I ask someone to imagine flying cars. Suppose further that this person has seen airplanes flying in the sky but has never seen cars flying like airplanes. In such cases, it seems reasonable to say that even if this person has never seen flying cars before, just from his exposure to flying airplanes, he/she is well equipped to make sense of the possibility of flying cars. As this analogy shows, the gap for this person has already been narrowed to make sense of the possibility of flying cars, despite having never seen such a thing before. Here the possibility I am talking about is physical possibility. In this case, for example, we can recombine the ideas of flying that we find in planes and being a vehicle, and substitute the plane shape for the car shape. How the details work out here, of course, is something that those with specialty in Aerospace Engineering can tell us about.

Of course, I am not implying here that every thought experiment has to be set up in a way I am suggesting. But as a general principle, thought experiments must in some sense either be grounded in experience or at least have some kind of proximity to our experience of the world. To make sense of all this, I will now briefly explain the three conditions that I had introduced at the beginning of this chapter. My goal in introducing these conditions is to show in what way thought experiments must be used in personal identity discussions. That is to say that, to get a maximum benefit from the use of thought experiments in personal identity discussions, at least the following three conditions must be satisfied, namely the Role Condition, the Constraint Condition, and the Priority Condition. That said, I must stress that in proposing these conditions, I am not attempting to provide necessary and sufficient conditions for the use of the methodology of thought experiments. Rather the conditions must be taken only as suggestions that I am making which, if practiced consistently, could help us think in a more effective way about the problem of personal identity.

1. The Role Condition. Given this condition, thought experiments in personal identity discussions must be used as illustrative devices. That means that we should resist the temptation of crafting thought experiments to make them yield the result we want. Rather, the metaphysical conclusion we reach about the nature of personal identity must be rooted in independent arguments we provide in support of it. In that case, we must make sure that we are using thought experiment only to help us clarify or add emphasis to our arguments (cf. Sorensen 1992: 15). If we grant this, then thought experiments must not be used to set the terms of the debate for personal identity discussions. In light of this, it remains unclear whether this is a kind of role Parfit's Teletransporter plays. My skepticism in this regard arises from the fact that once we accept Parfit's thought experiment, we will be forced to embrace a serious metaphysical assumption that Parfit advocates, which is that 'identity is not what matters in survival.' The question remains: Are we supposed to accept the metaphysical conclusion that identity is not what matters in survival because of the Teletransporter scenario or is there an independent reason that forces us to embrace it? It is not clear to me how in which of these senses we should understand the role of Parfit's Teletransporter (cf. Gendler 2002: 34-53; Baillie 1993: Ch. 5). As we shall see in chapter five,

the same question can be raised regarding other popular thought experiments in personal identity literature.

2. The Constraint Condition. Given this condition, thought experiments in personal identity discussions must be rooted in experience in one of the following senses: (i) directly grounded in experience, that is thought experiments that are physically possible in various circumstances; (ii) have some proximity to experience, that is to say that thought experiments that are indirectly related to a lived experience; and (iii) drawn from observations made based on actual cases. For present purposes, there is no need to advance a detailed discussion on (i)-(iii). One example here that best sums up (i)-(iii) is Edmund Gettier's famous thought experiment (1963). One thing that is distinctive about Gettier's case is that it is very much linked to what can actually happen, in fact often happens, to most us in real life. On many occasions we might have formed a false belief about a particular matter, without knowing that the belief we held at the time was false. Yet we were justified in those occasions until the real facts came to light, in light of which we had to revise our beliefs. So Gettier's case, is not something that is removed from our own experience. In light of this, the point of the Constraint Condition is that we should not conjure up any scenario we like in thinking about personal identity. Again seen in this way, Parfit's Teletransporter is far removed from anything we know in our experience. Hence, it is hard to see its merit in shaping in any positive way how we should think about personal identity. So as Brown and Fehige remark, "It seems right to demand that they also be visualized (or perhaps smelled, tasted, heard, touched); there should be something experimental about a thought experiment," (1996/2014).

3. The Priority Condition. Given this condition, thought experiments in personal identity discussions must not be given a chance to set the terms of the debate. What this means is that discussions about personal identity must begin with what I called the *Aristotelian Question*. In the literature on personal identity, it is common to see philosophers (as we saw in the case of Parfit), opening their discussion with science fiction puzzle cases and spend the rest of their time trying to work out their implications (see e.g., Perry 1972, 2002: Ch. 3; Salmon 2006: Ch.11). But this approach is deeply problematic, for all the reasons we saw in

the earlier sections. So in this thesis, I continue to insist that thought experiments must not be let to take the place of the *Aristotelian Question*. So the point of the Priority Condition is that in personal identity discussions, thought experiments must not be given the driver's seat.

Having discussed what thought experiments are and also how they should be used in personal identity discussions, I now turn to chapter five where, among other things, I will argue that the place of substance ontology for personal identity on the one hand, and the self/person on the other hand, should be reconsidered. However, as we shall see, unduly emphasis on thought experiments continues to eclipse the centrality of substance ontology for personal identity. In this regard, the problems I pointed out in this chapter with respect to the use of thought experiment again will surface in the way philosophers attempt to establish personal identity without making reference to substance ontology.

Chapter five

Substance and Substantial-Self/Person Question in Contemporary Personal identity Debates

As we recall, in chapter one in § 1.2, I identified the central issue in the contemporary personal identity debate as having to do with what Martin and Barresi described as *the traditional question*. This is the question that concerns itself with providing a necessary and sufficient condition for diachronic personal identity. Formally described, the traditional question can be stated in terms of the Lowe-Olson-Hudson schemata according to which, necessarily, if X is a human person at time t_1 and Y exists at another time t_2, $X=Y$, if and only if_____ (see again chapter one §§ 1.2.1 and 1.2.2). Since Locke first formulated the problem of diachronic personal identity (see e.g., chapter two), contemporary analytic philosophers' effort since the 1970s has been attempting to fill out the biconditional at the end of the above schemata. Here the most popular methodology philosophers employ is thought experiments. In this chapter, as we shall see, the methodology of thought experiments is closely related to solutions philosophers propose to the problem of the criterion of identity. In this case, the former often provides philosophers with a way to think about various sorts of scenarios (often drawn from science fiction) which are deemed helpful in establishing the criterion of identity. In light of this, our discussion of thought experiments in chapter four nicely fits into the present discussion.

In the first part of this chapter, my discussion will focus on the neo-Lockeans' response to the famous Butler-Reid objections against Locke's Memory Criterion. In the second part, I will mainly focus on the use of puzzle cases and their relation to the criterion of personal identity. In this chapter, my main goal is to argue that contemporary philosophers' undue emphasis on thought experiments and the criterion of personal identity has only contributed to the present stalemate in contemporary personal identity debates. Most of all, in this chapter, I will argue that the imbalanced emphasis put on the above two aspects is also to blame for the negligence of the proper subject matter in contemporary personal identity discussions. Here by proper subject matter, I mean the *'Aristotelian Question'* or

what I introduced in chapter one as the central question of this thesis, i.e., *the substantial-self/person question*. As we recall, *the substantial-self/person question* concerns with whether or not the self/person is a substance. My own answer to this question, as I explained in chapter one (see again § 1.1.2.3 and 1.1.2.4) is affirmative. But my defence of this affirmative answer will wait until I get to chapter seven. That said, ultimately, in this chapter I want to show why Shoemaker's observation about the contemporary philosophers' negligence of the *substantial-self/person question* is justified (see § 1.1). But here in saying that Shoemaker's observation is justified, I do not mean to condone such negligence itself, but only the accuracy of the observation. To remind ourselves once again, Shoemaker's observation was that nowadays the question of whether the self is a substance and whether personal identity over time requires a substance, has a musty smell to it. Despite the centrality of the question in discussions of personal identity in Locke, Butler, Hume, and Reid, it has not been the central question in contemporary discussions of personal identity. As I pointed out in chapter one, Shoemaker's observation is indicative of three distinct, but interrelated issues that need to be examined carefully. These are: (i) the rejection of the substantial-self for personal identity over time; (ii) the rejection of the centrality of *the substantial-self question* for personal identity; and (iii) the absence of *the substantial-self question* from most discussions of personal identity, respectively.

As we saw in chapter two, Locke paved the way for (i) in rejecting the substantiality of the self/person. In light of our discussion of chapter two, I would say that Locke was also responsible in paving the way for (ii) despite allowing substance to play a central role in his theory of personal identity. But many philosophers claim that it was Hume who put the final nail on the coffin when it comes to (i) and (ii) (see e.g., John Searle 2004: 192; Kolak and Martin 1991: 165-167; cf. Kripke 1980: 155 footnote # 77). So given (i)-(ii), most contemporary philosophers feel entitled to endorsing (iii) above. It is also at this point that most contemporary philosophers retreat from serious discussion of the ontology of the self while keeping themselves busy with what are sometimes extremely confusing discussions involving thought experiments and the criterion of identity. In this chapter, I will pursue my point that there is no good reason to endorse (i)-(iii) above.

I contend that we have no other more plausible option than taking the ontology of the self seriously. Ultimately the discussion in this chapter is meant to prepare us for chapter six, where I will attempt to show why the naturalistic ontology that attempts to undermine the realist ontology of the self is far from a live option to adopt. In that chapter, we will also see how it is possible to engage in a sensible discussion of the ontology of the self/person without necessarily relying on science fiction scenarios. That said, my subsequent discussion in this chapter will be located within the current most influential approach taken to the problem of personal identity. This is the neo-Lockean view, which focuses mainly on the Psychological Criterion. [89]

5.1 *The Psychological Criterion and Personal Identity*

The neo-Lockeans' notion of the Psychological Criterion is a modified version of Locke's Memory Criterion. As we saw in chapter two, for Locke, a person *P* is identical with himself/herself of a year ago or two weeks ago, etc., if he/she remembers the thoughts he/she had or the action he/she had done a week ago or a year ago, etc. This is often called 'experience-memory' for personal identity (Parfit 1984: 205). An 'experience memory' is a type of memory that is rooted in one's first-person perspective. Taken this way, when one is said to remember something, this should be understood to mean that one remembers one's own experiences and actions from one's own first-person point of view. Or as Noonan

_____

[89] For the contemporary neo-Lockeans psychological approach for personal identity over time, see e.g., Shoemaker (1970, 1984: 90, 1999, 1997); Perry (1972, 2002); Parfit (1971; 1984: Part III; 2012); Noonan (2003); Nozik (1981); Unger (1990: Ch. 5; 2000); Lewis (1976); Hudson (2001); Garret (1998); Johnston (1987). For the Physical/Somatic/Animalist approach, see e.g., Olson (1997; 2007); van Inwagen (1990: esp., Ch. 14); Snowdon in Christopher Gill (ed.) (1990: 83-107); Snowdon (2014); Ayers (1991: 278-292); William (1988); Mackie (1999); Wollheim (1984); Wiggins (1967; 1980); Williams (1956-7, 1970; 1973); Thomson in Dancy 1997: 202-229). For the constitution view of persons and personal identity over time, see Baker (2000: Ch. 5). For the simple view approach, see e.g., Chisholm (1976); Merricks (1998: 106-124); Lowe (1996, 2008); Swinburne in Shoemaker & Swinburne (1984). For a more recent discussion on the Simple view approach see, Gasser and Stefan, eds., (2012). For a transcendental-Kantian approach, see e.g., Doepke (1996). For an agency based approach, see e.g., Rovane (1998). For a narrative approach see, e.g., Schechtman (1996; also see 2014); also cf. MacIntyre (2007, 1981: esp. Ch. 15); Taylor (1989; esp. Ch. 2); Ricoeur, trans., Blamey (1992: esp. Chs. 5-6); Carr (1986). For important anthologies on personal identity, see e.g., Perry (1975); Rorty (1976); Kolak and Martin (1991); Noonan (1993); Gasser and Stefan (2012); Martin and Barresi, eds., (2003). For discussions that focus on various aspects of personal identity, see e.g., http://philpapers.org/browse/theories-of-personal-identity accessed May 10, 2014. For comprehensive bibliography on work on identity over time, see e.g., Sider (2000: 81-89). For a comprehensive reference on identity and personal identity, see Gendler (2000).

remarks, in experience-memory, one reports one's own experiences and actions in first person memory claims (1993: XIV). So in Locke's view, if a person P_2 at t_2 and a person P_1 at t_1 are the same, then P_2 is identical with P_1 if and only if P_2 and P_1 satisfy the memory condition as described above. However, not everyone embraces Locke's Memory Criterion for personal identity as stated above. In this regard, the objections of Locke's earliest critics, Butler and Reid take centre stage.[90] As we shall see, it is in light of the Butler-Reid objections that the neo-Lockeans modify Locke's Memory Criterion.

5.1.1. *The Butler-Reid* Objections and *the Core Assumptions* [91]

Although both Butler and Reid raised many objections[92] against Locke's Memory Criterion in the literature, two objections appear time and time again. Butler's often cited objection concerns with the problem of *circularity*. As Butler put it:

> But though consciousness of what is past does thus ascertain our personal identity to ourselves, yet, to say that it makes personal identity, or is necessary to our being the same persons, is to say, that a person has not existed a single moment, nor done one action, but what he can remember; indeed none but what he reflects upon. And one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes (In Perry 1975: 100).

Similarly, Reid's often cited objection concerns with the problem of *self-contradiction*. Reid claims that given Locke's criterion of personal identity, "*a man may be, and at the same time not be, the person that did a particular action,*" (in Perry 1975: 114; original emphasis). Here, Reid's point is that Locke's memory account for personal identity violates the transitivity relation—which is one of the properties of the logic of identity. Formally, it can be expressed as follows: $(\forall x)\,(\forall y)\,(\forall z)\,[(x=y \;\&\; y=z) \rightarrow (x=z)]$, (i.e., for all X, for all Y, for all Z,

---

[90] At this point, it may be asked why these objections have not already been discussed in chapter two, where Locke's theory of personal identity was discussed. I delayed this discussion until now because its relevance is more tied to the neo-Lockeans' attempt to modify Locke's Memory Criterion.

[91] Also see Leibniz's criticisms of Locke's memory criterion, see *The New Essays Concerning Understanding* (1981; see also Noonan 2003: Ch. 4).

[92] See e.g., Butler in Perry (1975: Ch. 5). This section is taken from Butler's the *First Dissertation to The Analogy of Religion*, London, MDCCXXXVI, (1736: 301-308). Also see Reid in Perry (1975: Chs. 6 and 7). This section is taken from chapter four of Ried's *Essays on the Intellectual Powers of Man* (1785).

if *X* is identical to *Y* and *Y* is identical to *Z*, then *X* is identical to *Z*). So how does Locke's memory account violate the transitivity relation? Reid explains it as follows:

> Suppose a brave officer to have been flogged when a boy at school for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life; suppose, also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that, when made a general, he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging (Ibid. 114).

In light of this example, Reid further remarks:

> From Mr. Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging; therefore, according to Mr. Locke's doctrine, he is not the person who was flogged. Therefore the general is, and at the same time is not, the same person with him who was flogged at school (Ibid. 114-115).

Before we look at the neo-Lockeans' response to the Butler-Reid Objections as stated above, it is important to keep in mind that both Butler and Reid embrace a substance-based account of personal identity. That is, both of them held that personal identity consists in the identity of substance (see e.g., Butler in Perry 1975: Ch. 5; Reid in Perry 1975: Chs. 6 and 7). In light of this, their conception of personal identity is rooted in three core assumptions:

(1) The identity condition of a self/person is different from that of the identity condition of other things, such as trees. For example, Butler describes the difference by drawing a distinction between 'strict and philosophical' sense of the term 'same' (i.e., identity)— applied to selves/persons on the one hand and 'a loose and popular' sense of the term 'same'—applied to other things such as trees, on the other hand. Butler argues that the identity we attribute to things like a tree should be taken in a loose sense, since such things constantly lose particles of matter that compose them. So Butler thought that at any given time, what we call a tree is not strictly speaking, a numerically one and the same tree that

persisted over time (in Perry 1975: 100-101).[93] Along similar lines, Reid also claims that the identity of persons is perfect. By that he means that it admits of no degrees (*Ibid*. 111). But the identity we attribute to bodies, whether natural or artificial, is not perfect, even though for convenience of speech, we call identity. In explaining why he denies fixed identity for bodies, Reid claims that except in the case of persons, in all other cases, identity "admits of a great change of subject, providing the change be gradual; sometime, even of a total change…questions about the identity of a body are very often about words," (*Ibid*. 112).

(2) Personal identity over time strongly presupposes numerical identity. For example, Butler claims that, "by reflecting upon that which is myself now, and that which was myself twenty years ago, I discern they are not two, but one and the same self," (*Ibid*. 100). Similarly, Reid insightfully remarks:

> My personal identity…implies the continued existence of that indivisible thing which I call *myself*. Whatever this self may be, it is something which thinks, and deliberates, and resolves, and acts, and suffers. I am not thought, I am not action, I am not feeling; I am something that thinks, and acts and suffers. My thoughts, and actions, and feelings, change every moment; they have no continued, but a successive, existence; but that *self*, or *I*, to which they belong, is permanent, and has the same relation to all the succeeding thoughts, actions, and feelings which I call mine (*Ibid*. 109).

(3) Personal identity is primitive in a sense that it is unanalysable in terms of more fundamental facts than itself. In other words, no necessary and sufficient conditions can be given for it. For example, Butler claims, "when it is asked wherein personal identity consists…all attempts to define, would but perplex it," (*Ibid*. 99). Similarly, Reid claims that, "if you ask a definition of identity, I confess I can give none; it is too simple a notion to admit of logical definition," (*Ibid*. 108).

_____

[93] Butler does not elaborate on these distinctions in detail. A well-known contemporary defense of these distinctions is due to Chisolm. For example, in talking about the 'loose sense of identity' Chisolm claims that, "familiar physical things such as trees, ships…may be construed as presupposing that these things are 'fictions', logical constructions or *entia per alio*. And it tells us that, from the fact that any such physical thing may be said to exist at a certain place P at a certain time *t* and also at a certain place Q at a certain other time *t'*, we may *not* infer that what exists at P at t is identical with what exists at Q at *t'*," (1976: 97). With respect to strict identity, Chisolm remarks, "this may be construed as telling us that persons are not thus 'fictions', logical constructions or *entia per alio*. And so it implies that, if a person may be said to exist at a certain place P at a certain time *t* and also at a certain place Q at a certain other time *t'*, then we *may* infer that something existing at P at *t* is identical with something existing at Q at *t'*," (*Ibid*.).

I shall call (1)-(3) above together the *Three Butler-Reid Assumptions*. At this point, one could ask: how would Locke on the one hand and his modern day followers, the neo-Lockeans, on the other hand would react to the *Three Butler-Reid Assumptions*? In light of our discussion in chapter two, Locke would have no problem of endorsing (1), since in his view, different sortal terms (e.g., man, persons, animals) are associated with different criterion of identity. But Locke would disagree with both Butler and Reid, when they claim that the loss of a material composition of a certain object, say of, a tree would make it impossible for such an object to maintain its identity over time. As we recall, Locke maintains that the identity of living things such as a tree consists in the same life they partake as opposed to, in the changing material particles that compose them and contribute to the overall internal organization of that life (see again chapter two, footnote # 64; see also Shoemaker in Shoemaker and Swinburne 1984: Ch. 14). Locke would also have no hesitation to embrace (2), since in his view, despite the changing bearers of persons (i.e., substances), numerically the same person is meant to persist over time. However, Locke clearly does not endorse (3), since in his view, personal identity consists in the identity of consciousness. Yet regardless of such a key difference, as the hitherto brief discussion indicates, Locke seems to have some common ground with Butler's and Reid's conception of personal identity.

On the other hand, the neo-Lockeans, as we shall see, are very much critical of the *Three Butler-Reid Assumptions*. To see this, we need to recall the distinction I discussed in chapter one between the complex views and the simple view (see § 1.2.2 C). As we recall, views that belong to the former category reject the notion that the identity of persons is different from the identity of other things, say of, trees. That means that for defenders of the complex view, a non-circular, non-trivial and informative constitutive criteria can be established for personal identity (see e.g., Shoemaker in Gasser and Stefan 2012: Ch. 6; see also Noonan *Ibid*. Ch. 4). In this case, since the neo-Lockeans are defenders of the complex view, they do not embrace (1). Their rejection of (1) in turn, paves the way for their rejection of (3) as well. This is because, since according to the neo-Lockeans' view, the identity condition of persons is not different from the identity condition of other things, personal identity is not a primitive or unanalysable fact.

Where does this leave (2) then? The neo-Lockeans do not take (2) at face value. Here a good case in point is Parfit's famous argument that numerical identity is not what matters in survival (see e.g., 1984: Ch. 12). Along similar lines, Perry also argues that "neither identity, nor the belief in identity, nor even the imagining of identity, is necessary for identification," (in Rorty 1976: 77). For reasons we shall see, both Parfit's as well as Perry's remarks about identity are far from convincing. That said, however, in rejecting to embrace (2) at face value, the neo-Lockeans, such as Parfit, Perry and others, are not denying Lewis's dictum about self-identity, i.e., 'nothing fails to be identical to itself' (1986: 192-93). Rather, the neo-Lockeans' point is that in certain puzzle cases (2) suffers from a problem of vagueness/indeterminacy. Here by 'indeterminacy,' it is meant to refer to states of affairs, such as "whether a certain object has or lacks a certain property. If neither of these is the case, then the state of affairs is indeterminate, and a sentence reporting it lacks truth-value," (Parsons 2000: X). For example, Perry illustrates such vagueness in personal identity by appealing to a puzzle case he calls, ''brain-rejuvenation''. As Perry puts it:

> Smith's brain is diseased; a healthy duplicate of it is made, and put into Smith's head. On the assumptions about the role of the brain usually made in these discussions, the survivor of this process will be just like a healthy Smith. But will he be Smith? It seems that people of good faith can differ over the answer to this question (and, if this case is not one that is truly indeterminate, it points in the direction in which such a case could be constructed) (in Rorty 1976: 68; for details see Perry 1972: 463-488).

In light of such puzzle cases, the neo-Lockeans argue that, even if there is no indeterminacy in identity *simpliciter*, personal identity can be indeterminate (see e.g., Noonan in Gasser and Stefan 2012: 95-101). Moreover, personal identity theorists also defend the notion of vagueness in personal identity by responding to Evans's famous argument against the idea of vague objects (see e.g., Evans 1978, Garrett 1998: Ch. 5; Tye 2003: 154-163). As Sorensen nicely sums up, the gist of Evans's argument runs as follows:

> If there is a vague object, then some statement of the form '$a = b$' must be vague (where each of the flanking singular terms precisely designates that object). For the vagueness is allegedly due to the object rather than its representation. But any statement of form '$a = a$' is definitely true. Consequently, a has the property of being definitely identical to a. Since $a = b$, then b must also have the property of being definitely identical to a. Therefore '$a = b$' must be definitely true!" (Sorensen 2012: § 8).

But the neo-Lockeans could not accept Evans's argument for the simple reason that given their relational view of personal identity, vagueness in personal identity is unavoidable. However, this will not be true if one approaches the issue of personal identity from a non neo-Lockean standpoint. This is what I will attempt to do in chapters seven and eight. That said, as we shall see in § 5.1.2, of the *Three Butler-Reid Assumptions*, the one that takes centre stage in the neo-Lockeans' attempt to modify Locke's Memory Criterion is (2). This is because, the two Butler-Reid Objections, namely the circularity and the self-contradiction (or 'Brave Officer Paradox')[94] have direct bearing on (2).

5.1.2 *The Neo-Lockeans' Response* to *the Butler-Reid Objections*

The neo-Lockeans' strategy in responding to the Butler-Reid Objections can be analogously explained using Strawson's famous distinction in his *Individuals*, between two approaches he calls: 'descriptive metaphysics' and 'revisionary metaphysics.' In the case of the former, Strawson tells us that one engages in describing the actual structure of our thought about the world. In the case of the latter, one's aim is to produce a better structure (1959: 9). Similarly, the neo-Lockeans strategy in responding to the Butler-Reid Objections has been revisionary. That is to say that the neo-Lockeans' focus is not on describing Locke's Memory Criterion in its own terms, whether or not it faces the Butler-Reid type objections. Rather, as we shall see, what the neo-Lockeans strive to do is to revise Locke's account of the Memory Criterion in such a way that it escapes the Butler-Reid type of objections. By taking such a revisionary manoeuvre, the neo-Lockeans may well be able to show that Locke's theory of personal identity could escape or resist Butler-Reid type objections. Yet, as we shall see, it remains unclear whether or not Locke himself would endorse the neo-Lockeans' revisionary analysis of his view of the Memory Criterion. For reasons we shall see, my own view is that Locke would have some serious hesitations to embrace the neo-Lockeans' revisionary analysis of his Memory Criterion.

Having said that, one of the ways the neo-Lockeans apply their revisionary analysis to

---

[94] The phrase 'Brave Officer Paradox' is commonly used in the literature (see e.g., Perry 2002: 86).

Locke's Memory Criterion is by divesting (or in some way weakening) the first personal aspect of Locke's experience memory. As indicated in § 5.1, by its very nature, 'experience memory' is first personal. That is to say that 'experience-memory' is often reported in first person memory claims (e.g., I remember breaking my leg two years ago). In the literature, 'experience-memory' also goes by other labels, such as 'episodic memory', 'biographical memory', 'direct memory' and 'personal memory' (see e.g., Sutton 2010: § 1). This form of memory is distinguished from other forms of memory, such as (i) 'procedural memory,' which concerns with skill memory (e.g., I remember how to drive a car); and (ii) 'propositional memory', which concerns with memory of facts (e.g., I remember that St. Augustine lived in 4th century A.D) (see further Sutton 2010). Among notable contemporary neo-Lockeans, who revised Locke's Memory Criterion include, Grice (1941/1975), Quinton (1962/1975), Perry (1972, 2002; 1975), Shoemaker (1970, 2003, 1984: Part Two), Noonan (1989/2003), Parfit (1971; 1984); and Garrett (1998).

As far as the neo-Lockeans' response to the Butler-Reid Objections is concerned, Parfit's revisionism is the most influential. So my subsequent discussion in this regard will focus mainly on Parfit's approach. However, my goal, here, is not just to spell out the neo-Lockeans' position for its own sake. Rather my main interest is to reiterate the point I have been making up to now regarding the marginalization of *the substantial-self/person question*. As I tried to show in the previous chapters why there are no good reasons that justify the marginalization of *the substantial-self/person question*, I will advance a similar argument in discussing the neo-Lockeans' view.

A. *Parfit's Revisionism*

Parfit modifies Locke's Memory Criterion by appealing to the concept of an overlapping chain of experience-memories. In Locke's view, as we recall, what allows a person's identity to hold over time is what Parfit calls, '*direct memory connections*' (1984: 205; emphasis original). To use Parfit's own example, we can say that there are direct memory connections between *X* today and *Y* twenty years ago, if *X* can now remember some of the experiences of *Y* twenty years ago. Taken this way, whenever there is a break in direct memory connections, given Locke's theory, a person cannot persist over time. But as Parfit rightly

points out, if Locke's view in this regard was true, then it would not be possible for a person to forget any of the things he/she once did or any of the experiences that he/she once had (*Ibid*.) But this seems to fly in the face of our experience, since we often forget the things that we did or experience in the past. But from this, it hardly seems to follow that we no longer exist as persons.[95] In this case, the neo-Lockeans understandably claim that they have a good reason in seeking to amend Locke's Memory Criterion. In this regard, Parfit's initial revisionary step concerns with substituting Locke's direct memory connection requirement, for what he calls, ''*continuity of memory*''—i.e., between X now and Y twenty years ago,'' (*Ibid*. emphasis original). One of the advantages of such a revised version of Memory Criterion is that it allows a person's memory to have some gaps. This in turn would alleviate the worry Locke's strict memory connection requirement raises for a person's persistence over time (for reasons we already saw). Given Parfit's modified 'continuity of memory' requirement, a person may not be able to remember some of the things he/she did or experience in the past. But as Parfit explains, "on the revised Locke's view, some present person X is the same as some past person Y if there is between them continuity of memory," (*Ibid.*). For Parfit, this revision is just one small step in the right direction. To make Locke's view more defensible, additional revision must be done. In this case, Parfit remarks:

> We should…revise the view so that it appeals to other facts. Besides direct memories, there are several other kinds of direct psychological connections. One such connection is that which holds between an intention and the later act in which this intention is carried out. Other such direct connections are those which hold when a belief, or a desire, or any psychological feature, continues to be had (*Ibid*. 205-206).

Here Parfit's point is that, since the domain of direct experience-memory is limited, we can

---

[95] In some sense, the experience-memory constraint that Locke puts on a person's continued existence seems to echo Galen Strawson's, the *Pearle view* of selves/persons. According to Strawson's Pearle view, "a mental self exists at any given moment of consciousness or during, any uninterrupted or hiatus-free period of consciousness," (in Martin and Barresi 2003: 360). But no less than Locke's experience-memory view, Strawson's Pearle view is also unsatisfying. This is because, there are many occasions a person may slip in and out of consciousness. In this case, an everyday 'sleep' routine is one of those occasions where a person becomes temporarily unconscious. Moreover, it is not uncommon to see because of an injury or medically induced situation, a person might become unconscious for a short or long period of time and regains his/her consciousness. In all such cases, it is perfectly plausible to say that a person who becomes temporarily unconscious is the very same person with a person who later regains his/her consciousness. This is the view that common-sense is on its side or so it seems to me.

expand it by incorporating psychological relations that hold between various mental states. In light of this, Parfit specifies the nature of the psychological relations he has in mind by drawing the following key distinction (*Ibid*. 206):

> *(A) Psychological connectedness* is the holding of particular direct psychological connections (emphasis original).

> *(B) Psychological continuity* is the holding of overlapping chains of *strong* connectedness (emphasis original).

The reason I described (A) and (B) as a key distinction is because it underlies the neo-Lockeans' response to the Butler-Reid Objections. To see this, we need to recall the remark I made earlier regarding the nature of the neo-Lockeans' revisionary analysis of Locke's Memory Criterion. In this case, my remark was that one of the ways that the neo-Lockeans apply their revisionary analysis to Locke's Memory Criterion is by divesting (or in some way weakening) the first personal aspect of Locke's experience memory. The question remains: How does Parfit's distinction in (A) and (B) above help the neo-Lockeans achieve that goal? To properly answer this question, first it is important to understand Parfit's analysis of (A) and (B). Parfit claims that, of the two relations described in (A) and (B), connectedness is more important from the standpoint of both theory and practice. For Parfit, connectedness is a matter of degree. This means that between X at t_2 and Y at t_1, there might be thousands of direct psychological connections. What is needed for X at t_2 and Y at t_1 to be the same person, is the presence of enough direct psychological connections between them. Here one may ask, what counts as 'enough'. Parfit's response to this question is that, since connectedness is a matter of degree, no precise definition can be given for what counts as 'enough'. For Parfit, when there are enough direct psychological connections, there is what he calls '*strong* connectedness,' (*Ibid*. 206).

Notice that in (B), Parfit defines psychological continuity in terms of the notion of 'strong connectedness.' Put differently, Parfit collapsed (A) into (B), by stripping away from the former, the requirement for direct psychological connectedness. In doing so, Parfit seeks to show how the notion of continuity can be integrated with the notion of connectedness (cf. Noonan 2003: 54-55). Put this way, (B) allows a person to have some strong psychological

connections with his/her past, without at the same time, requiring such connections to be strong, all the way into the distant past. That is, (B) allows for some memory gaps in a person's life, without putting his/her identity at stake. In this case, all that a person needs to have is strong connection to his/her immediate past. For example, following Parfit, I can say that, I am now strongly connected to myself a month ago, when I was strongly connected two months ago and so on. From this, it does not follow that I am strongly connected to myself fifteen years ago. As Parfit remarks,

> Between me now and myself twenty years ago there are many fewer than the number of direct psychological connections that hold over any day in the lives of nearly all adults. For example, while these adults have many memories of experiences that they had in the previous day, I have few memories of experiences that I had twenty years ago (*Ibid*.).

In addition to allowing memory gaps in a person's life, Parfit's notion of strong connectedness as discussed so far, seems to have another advantage for the neo-Lockeans. For example, Parfit claims that since strong connectedness is a matter of degree, it is not a transitive relation. Since it is not a transitive relation, it cannot be the criterion of identity (*Ibid*.). At this point, the advantage for the neo-Lockeans seems to be that strong connectedness can be given as a response to Reid's the Brave Officer Paradox. As we recall, Reid's main objection against Locke's criterion of personal identity concerns with its violation of the transitivity relation. By adopting Parfit's notion of strong connectedness, the neo-Lockeans can claim that Reid's objection can be diffused. For example, Noonan remarks:

> What is needed is just the distinction between *connectedness* of consciousness and *continuity* of consciousness, where continuity is defined in terms of connectedness by saying that a later person *P*2 at *t*2 has a consciousness which is continuous with that of an earlier person *P*1 at *t*1 just in case he is the last link in a chain of connecting persons beginning with *P*1 at *t*1, each of whom is conscious of the experiences and actions of the preceding link in the chain…a revision of Locke's account which makes personal identity consist in continuity of consciousness is immune to…objections of Butler and Reid (Noonan 2003: 54-55; see also Perry 1975: Ch. 1).

So based on Parfit's account of strong connectedness, the neo-Lockeans might respond to Reid's the Brave Officer Paradox, by saying that an officer who was flogged at school is strongly connected with the officer who took the standard now. The officer who took the

standard is strongly connected with the same person who was made a general. But the officer is not strongly connected to the same person who was flogged at school as a child. It does not follow that the general is strongly connected to the same person who was flogged at school. But despite the presence of such a memory gap, the officer remains identical with the person who was flogged at school as a child. Contra Reid, the neo-Lockeans may conclude that there seems to be no violation of the transitivity relation. So Reid's self-contradiction objection is non-existent.

How does the hitherto Parfit's revised version of Locke's view fair against Butler's circularity objection? Parfit claims that the revised version does not answer Butler's objection. To see why this is the case, Parfit reminds us that, "The Psychological Criterion appeals, not to single memories, but to the continuity of memory, and, more broadly, to Relation R, which includes other kinds of psychological continuity," (1984: 119). So Parfit claims that, on one interpretation, Butler's objection could be put as follows:

> It is part of our concept of memory that we can remember only *our* own experiences. The continuity of memory therefore presupposes personal identity. The same is true of your Relation R. You claim that personal identity just consists in the holding of Relation R. This must be false if Relation R itself presupposes personal identity (*Ibid*. 220).

Parfit answers this objection by using a technical notion, *quasi-memory*. This is the notion first introduced by Shoemaker, who states it as follows:

> We need to consider…whether there could be a kind of knowledge of past events such that someone's having this sort of knowledge of an event does involve there being a correspondence between his present cognitive state and a past cognitive and sensory state that was of the event, but such that this correspondence, although otherwise just like that which exists in memory, does not necessarily involve that past state's having been a state of the very same person who subsequently has the knowledge. Let us call such knowledge, supposing for the moment that it is possible, as "quasi-memory knowledge," and let us say that a person who has this sort of knowledge of a past event "quasi-remembers" that event (2003: 24).

Here the point is that quasi-remembering something leaves the identity of quasi-rememberer unspecified. For example, Shoemaker characterizes the difference between quasi-remembering and remembering as follows, "whereas someone's claim to remember a past event implies that he himself was aware of the event at the time of its occurrence, the

claim to quasi-remember a past event implies only that someone or other was aware of it," (*Ibid*.). Parfit also remarks:

> I have an accurate quasi-memory of a past experience if (1) I seem to remember having an experience, (2) someone did have this experience, and (3) my apparent memory is causally dependent, in the right kind of way, on the past experience. On this definition, ordinary memories are a sub-class of quasi-memories. They are quasi-memories of our own past experience (1984: 220).

Although it is not part of our normal experience to quasi-remember other people's past experiences, Parfit argues for its possibility by appealing to neuro-surgeons at some point being able to develop ways to create in one brain a copy of a memory-trace in another brain. Parfit says that this might enable us to quasi-remember other people's past experiences (*Ibid*.). In responding to Butler's objection, Parfit remarks:

> My mental life consists of a series of varied experiences. These include countless quasi-memories and these earlier experiences. The connections between these quasi-memories and these earlier experiences overlap like the strands in a rope. There is *strong connectedness* of quasi-memory, if, over each day, the number of direct quasi-memory connections is at least half the number in most actual lives. Overlapping strands of strong connectedness provide *continuity of quasi-memory…*Since the continuity of quasi-memory does not presuppose personal identity, it may be part of what constitutes personal identity (*Ibid*. 222). [96]

B. *Locke's Response to the Neo-Lockeans*

It would be helpful at this point, to step back and look at how Locke himself might have thought about the neo-Lockeans' revisionism of his Memory Criterion. I already conceded

_____

[96] But as Oderberg argues, it remains unclear whether quasi-memory is any less identity-presupposing notion compared to experience-memory, which it seeks to replace. For example, Oderberg, asks us to suppose that Jones is trying to establish whether it was smith he saw in Cornmarket two days ago. Jones tells Smith that he saw someone exactly resembles him walking down the street. At this point, Smith reaches out to his bank of quasi-memories to decide whether it was him whom Jones saw. Oderberg further asks us to suppose that Smith finds he has an apparent memory quasi-memory of his walking down Cornmarket at the specified time. Again Oderberg asks us to suppose that Smith is the person Jones saw. In this case, it is more reasonable to assume that Smith is the same person as the person Jones saw. This is because, among other things, Smith quasi-remembers doing what Jones claims to have seen him doing. Oderberg claims that quasi-memory might not presuppose the identity of the person who had the experience and also might not presuppose the identity of the person who has the apparent memory of it. However, as Oderberg points out, quasi-memory is still identity presupposing in the sense that, when there is identity, say of, as in the standard case, quasi-memory obtains precisely because there is identity (1993: 181-182). In light of such considerations, Parfit's response to Butler's circularity objection is far from decisive, to say the least. For other similar critical discussions on quasi-memory, see Wiggins (2001: 212-225); Hamilton (2013: Ch. 2). For those who defend the notion of quasi-memory, see e.g., Roache (2006).

that the neo-Lockeans' revisionism might be one of the ways to show how Locke's problematic Memory Criterion might be improved. But given my discussion of Locke's view in chapter two, I have some serious doubts whether Locke would ever happily embrace his modern day followers, the neo-Lockeans' revisionism of his Memory Criterion. We saw that the neo-Lockeans' revisionism mainly targets on stripping away the first personal aspect of experience-memory. But Locke would never have settled for such a manoeuvre in order to save his view from the Butler-Reid type objections. This is because, for Locke, personal identity is first and foremost a practical matter; and only derivatively theoretical. This fact makes Locke's approach to personal identity to be significantly different from his modern-day followers, the neo-Lockeans' approach. For the latter, personal identity is first and foremost a theoretical enterprise and derivatively a practical matter. For reasons we shall see, I even doubt that for some neo-Lockeans such as Parfit, personal identity is a practical matter at all.

That said, for Locke, personal identity has to answer one fundamental question. That is: How will God's distribution of reward or punishment in the future eschatological era be just? One receives just reward or punishment from God, if and only if, His justice is distributed to the right person. By the 'right person,' I mean the very same person who either deserves divine punishment or reward based on his/her own actions. But for this to happen, one must be able to remember what one had done or had experienced in the past. So Locke would not have embraced the neo-Lockeans' suggestion that, numerical identity can be made to collapse into mere psychological continuity, without thereby assuming the very same person persisting all the way through (see e.g., Parfit 1984: Ch. 12). Such proposal would have sounded very strange to Locke's ears. That is to say that Locke would not have entertained such a proposal seriously. If the observations I made in chapter two about Locke's core view of personal identity are correct, then the neo-Lockeans are mistaken in thinking that their revisionism puts Locke's view of Memory Criterion in a better light. I disagree! The truth is quite the contrary. As the hitherto discussion shows, the neo-Lockeans and Locke stand far apart from each other when it comes to the necessity of remaining numerically the same person. For Locke, no one can replace a person's identity. Moreover, for Locke, a Memory Criterion, is a cornerstone of his view of personal identity.

163

For example, Locke perfectly realizes that there could be a memory loss even beyond any hope of retrieving it. When that happens, Locke's own preference is not to revise his Memory Criterion. Instead, Locke adamantly continues in maintaining that, in the case of a memory loss, the first person pronoun "I" will no longer refer to the same person; but only to the same man (see again chapter two). This shows that Locke is extremely serious when it comes to the Memory Criterion. A person's identity necessarily consists in experience-memory. Since we know that Locke's Memory Criterion is problematic, what are we supposed to do then? My own suggestion is that we should allow Locke's theory of person as well as his Memory Criterion to fail on its own ground (cf. Lowe 2005: 87-97). That is, what I attempted to do in chapter two. So contra the overwhelming majority of contemporary personal identity theorists,[97] I hold that the Butler-Reid Objections are effective against Locke's Memory Criterion. The neo-Lockeans cannot just assume that their modified version of Locke's view, even if, granting that it is successful against the Butler-Reid Objections, can be brought in to Locke's aid.

If the neo-Lockeans' revisionism does not succeed in impressing Locke, does it succeed on its own ground as a viable theory of personal identity? This is the question I intend to explore in the next section.

5.1.3 *Puzzle Cases, Criterion of Personal Identity and the Substantial-Self/Person Question*

I believe that our discussion in §§ 5.1; 5.1.1 and 5.1.2 has given us a good understanding of the backdrop against which the neo-Lockeans modified Locke's Memory Criterion. In this regard, both Butler's and Reid's objections played a key role in terms of forcing the neo-Lockeans to come up with their own solutions. However, for reasons already given, the neo-Lockeans' solutions would not be acceptable for Locke. That said, my subsequent discussion will focus on the use of the puzzle cases in personal identity and the criterion of personal identity. In this section, I will also discuss how the contemporary personal identity theorists' emphasis on these two aspects have contributed to the marginalization of what I called the

---

[97] Here I have in mind mainly the neo-Lockeans such as Shoemaker (1970, 2003, 1984); Perry (2002); Noonan (2003); Garrett (1998) and Parfit (1984).

Aristotelian Question, i.e., *the substantial-self/person question*. I will approach all of these issues from the standpoint of the neo-Lockean view.

A. Criterion of Personal Identity and the Substantial-Self/Person Question

Disagreements over the criterion of personal identity remain unresolved. In this case, the debate is often dominated by two rival theories, namely: the neo-Lockeans' Psychological Criterion and the Physical Criterion (see e.g., Sider 2002b: 81-89; Mackie 1999: 369-376; Noonan 1998: 302-318; 2001: 83-90; Parfit 2012). Even though these two theories are rivals, as we shall see, they come together in rejecting *the substantial-self/person question*. To see this, we should look at the precise characterization of each of these views. Although by now, we have a good grasp of what the main features of the Psychological Criterion are, it would be helpful to formally contrast it with the Physical Criterion. Here, I follow Parfit's characterization of each of these criteria. Parfit describes them as follows:

> (A) *The Psychological Criterion*: (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) there does not exist a different person who is also psychologically continuous with Y. (5) Personal identity over time just consists in the holding of facts like (2) to (4).

> (B) *The Physical Criterion*: (1) What is necessary is not the continued existence of the whole body, but the continued existence of *enough* of the brain to be the brain of a living person. X today is one and the same person as Y at some past time if and only if (2) enough of Y's brain continued to exist, and is now X's brain, and (3) there does not exist a different person who also has enough of Y's brain. (4) Personal Identity over time just consists in the holding of facts like (2) and (3).

Parfit also tells us that both (A) and (B) are reductionist views. For Parfit, these views are reductionist because they claim:

> (i) that the fact of a person's identity over time just consists in the holding of certain more particular facts.

> (ii) that these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an impersonal way (1984: 210).

In defence of (ii), Parfit remarks:

When we describe the psychological continuity that unifies some person's mental life, we must mention this person, and many other people, in describing the *content* of many thoughts, desires, intentions, and other mental states. But mentioning this person in this way does not involve either asserting that these mental states are had by this person, or asserting that this person exists. A similar claim applies to the Physical Criterion (*Ibid*.).

Given Parfit's characterization of (A) and (B) on the one hand, and the reasons he states in (i) and (ii) on the other hand, there seems to be no room left for *the substantial-self/person question*. Parfit's defence of (ii) above substantiates my observation. Here the tone of Parfit's remarks is Humean in nature. As we saw in chapter three, Hume argued that perceptions have no owners, i.e., distinct bearers. Similarly, Parfit's defence of (ii) also echoes Hume's rejection of a distinct bearer of experience (s)" (*Ibid*. 216). Parfit claims that, views that reject either or both of (A) and (B) are non-reductionist. In the non-reductionist view, as Parfit points out, a person is distinct from his body and brain, and his experiences (*Ibid*. 210). Parfit rejects this view out of hand. As Parfit remarks, "we are not separately existing entities, apart from our brains and bodies, and various interrelated physical and mental events," (*Ibid*. 216). More specifically, in his most recent essay entitled, 'We are not Human Beings,' Parfit defended what he calls, 'Embodied Person View,' according to which, persons are a thinking part of the brain (see e.g., 2012: 17). For reasons we shall see in chapters six, seven and eight, I find Parfit's claim that persons are a thinking part of their brains as deeply problematic. However, Parfit's official rejection of the substantial view of the self remains the same (see 1984: Ch. 13). In this case, Parfit speaks for many personal identity theorists. That said, it is important to understand that the neo-Lockeans rejection of *the substantial-self/person question*, is mainly rooted in (A). Central to (A) is the idea that psychological continuity requires the 'right kind of cause.'

Parfit tells us that there are three different versions of the 'right kind of cause.' These are: the Narrow, the Wide and the Widest versions, respectively. In the case of the narrow version, the right kind of cause is taken to be the normal cause (*Ibid*). We can illustrate this point by using what Unger calls, the two aspects of a person's dispositional psychology (1990: 67). For Unger, dispositional psychology comprises of a person's core psychology and her distinctive psychology. Unger claims that among the mental capacities that a person's

dispositional psychology includes, we find those we share with all other normal human beings, namely our capacity for conscious experience, our capacity to reason and the capacity to form intentions, etc. Unger calls such capacities instances of a person's core psychology. On the other hand, Unger also claims that there are certain aspects of a person's psychology that he/she shares with some normal humans but do not share with others. An example of the latter could be a person's memory of having tasted vanilla ice cream. Unger calls this a person's distinctive psychology (*Ibid*. 68).

In light of such considerations, Unger further claims that a person's persistence condition is contingent (dependent) on what happens to these aspects of our psychology. Moreover, a person's persistence condition is also contingent on what happens to what realizes both the core psychology as well as distinctive psychology. In Unger's view, in the normal course of events, the realizer of the dispositional psychology is a person's brain. Unger claims that on the physical approach, the importance of a person's brain is entirely due to its realization of his/her core psychology (*Ibid*. 69). In Unger's view, once the physical realization is in place, the psychological approach may be taken to guarantee what is needed for a person's existence at a future time. That is, a person's present psychology be causally carried forward in time as a single future of his/her own psychology. So a person to be himself/herself, he/she at the later time should have both his/her core psychology as well as distinctive psychology (*Ibid*.).

On the other hand, in the case of the Wide version, the cause is taken to be any reliable cause. On the Widest version, the cause is taken to be any cause. An example of both of these two causes would be something like Parfit's teletransporter, which produces one's replica, who then becomes psychologically continuous with oneself (*Ibid*. 209). Parfit claims that on both the Physical Criterion and Narrow Psychological Criterion, one's replica would not be oneself. By contrast, on the two wide criteria, Parfit says that one's replica would be oneself (1984: 209). Notice that in each of the three versions above, a distinct bearer (i.e., mental entity) of psychological properties is ruled out. For example, in Unger's account, the brain is the bearer of a person's dispositional psychology (i.e., both the core psychology as well as distinctive psychology). But what remains unclear in Unger's account has to do with

what exactly personhood amounts to. That is, is a person identical with a brain or is it the sum of the psychological properties that the brain is said to realize? Whichever way one answers this question, Unger's account has nothing to say about *the substantial-self/person question*. Similarly, since Parfit's the Wide and the Widest versions would allow for psychological continuity to hold through an impersonal means such as teletransporter, in this case, the question of a substantial self does not even arise.

So both in Unger's and Parfit's accounts, *the substantial-self/person question* receives a negative answer. That means that the two most influential views of the reductionist tradition (i.e., the neo-Lockean approach and the physical approach) do not embrace the view that the self/person is a substance. In light of this, Olson is deeply mistaken when he claims that, "anyone who assumes *person* is a substance concept is in effect assuming the Psychological Approach," (1997: 29). However, it could be said that some neo-Lockeans such as Shoemaker argue that a person is a substance (see e.g., Shoemaker 2003: Ch. 17). But it is far from clear what to make of Shoemaker's own view in this regard. One of the difficulties we face in understanding Shoemaker's position on this issue has to do with his attempt to integrate his materialist position with his neo-Lockean view of a person (see e.g., in Gasser and Stefan 2012: Ch. 6; 1984: Part Two; also cf. 1963: Ch. 2). But this move has been criticized by some philosophers on grounds that materialism is incompatible with the Psychological Approach (see e.g., Van Inwagen 1997: 305-319). In § E, I will raise my own objection against the neo-Lockeans' notion of a person.

B. Criterion of Personal Identity, Puzzle Cases and the Current Debate

In one of the classic anthologies of the 1970s entitled, *The Identities of Persons,* Rorty nicely sums up the contemporary debate on personal identity as follows:

> Disagreement about the criteria for personal identity have been persistently unresolved, the battle lines repetitively drawn over the same terrain, along familiar terrain, along familiar geographical strongholds. Although they disagree among themselves about its analysis, defenders of a physical or a spatio-temporal criterion are ranged against defenders of a psychological criterion, themselves uneasily allied. Peacemakers who argue that neither the psychological nor the physical criterion can be applied without implicitly reintroducing the other have been drawn into the battle—as peacemakers often are—as third or fourth parties. Although the

controversy has a long history, and although many arguments have been refined by repeated firings, there is little reason to expect a resolution that will not in time lead to renewed hostilities. What is required is not more ingenuity for more elaborate strategies, but an understanding of the conflicting interpretations of what has been at issue….Also at issue are methodological disagreements about what is involved in giving a criterial analysis (1976: 1-3).

Three things stand out from Rorty's remarks above: (i) disagreements besetting the criteria for personal identity are often recycled. That is to say that arguments often appear in new clothes but with no substantial change in their content; (ii) even when progress is being made in the personal identity debate, it often gets hampered by hostile reception; and (iii) multiplying ingenious strategies in the personal identity debate does not necessarily result in genuine progress. Rorty made these remarks nearly forty years ago now. At this point, the question that naturally comes to mind is this: What progress has been made since then? Before answering this question, let us look further at Rorty's other remarks concerning the state of the contemporary personal identity debate.

Rorty claims that controversies about personal identity have been intensified owing to the fact that there are a number of distinct questions that have not always been clearly distinguished from one another. Rorty claims that those who engage in debate over these questions often fail to make clear which of these questions is centrally interesting. Here are the questions Rorty talks about (*Ibid*. 1-2):

(1) Class differentiation: What marks out the class of persons from their nearest neighbours, from baboons, robots, human corpses, corporations?
(2) Individual differentiation: What are the criteria for the numerical distinctness of persons who share the same general description? This can be taken as "the problem of individuation."
(3) Individual reidentification: What are the criteria for reidentifying the same individual in different contexts, under different descriptions, or at different times? The majority philosophers who focus on individual reidentification analyze conditions for temporal reidentification, trying to define conditions for marking out successive stages of a continuing person from stages of a successor or descendent person.
(4) Individual identification: What sorts of characteristics identify a person as *essentially* the person he/she is, such that if those characteristics were changed, he/she would be a different person, although he/she might still be differentiated and reidentified as the same?

The questions in (1)-(4), yield different answers depending on one's approach in dealing with them. By 'one's approach,' I mean whether one is trying to answer (1)-(4) within the framework, say of, animalism, neo-Lockeanism, constitution view of persons, substance view of persons and so on. To see this, first it is important to understand the key question that echoes at the background of (1)-(4). The question I have in mind is this: How do persons relate to their bodies? Let us call this the *background question*. One's answer to the background question constrains how one tackles the questions in (1)-(4). Thomson once remarked that, "the simplest view of what people are is that they are their bodies," (in Dancy 1997: 202). Along similar lines, Williams also argues that persons are 'material bodies,' (1973: Ch.5). If one settles for Thomson and Williams's answer to the *background question*, then no further questions can be raised about it. But in reality, that is not the case. For example, some philosophers take the *background question* as a puzzle. In his *Indeterminate Identity*, Parsons lists four puzzles, among which, two of them have to do with the notion of a person. These are: (a) Is a person identical with that person's body?; and (b) If a person undergoes a crucial change, is the person after the change identical with the person before the change? (2000: X).

Notice that what I called the *background question* is directly linked to Parsons's (a). In fact, (a) can be taken as a paraphrased version of the *background question*. Parsons thinks that since it is indeterminate (i.e., there is no fact of the matter) whether or not a person is identical with that person's body, there is no solution for (a). Similarly, Parsons claims that since it is indeterminate whether or not a person retains its identity through drastic personality change, we cannot say anything definitively with respect to (b). Whether or not we agree with Parsons's own conclusions with respect to (a) and (b), the issues raised here are central to our conception of personhood. This is because, in answering (a), we come to understand what kind of entities persons fundamentally are. On the other hand, in answering (b), we come to grasp the persistence condition of persons—i.e., what kind of changes persons can and cannot survive. Notice that (a) and (b) are intimately related in that to get a grip on (b), first we need to get a grip on (a). Taken this way, (a) and (b) inform each other. In the contemporary personal identity debate, (a) and (b) are overshadowed because of the priority given to the question of the criterion of personal identity. This is

nowhere more evident than in the neo-Lockean tradition. Before taking up that discussion, I want to briefly point out some of the non neo-Lockeans' responses to (a) and (b).

For example, if it is seen from the animalists' standpoint, (a) should be answered affirmatively. Noonan describes animalism as the "thesis that any person coincident at any time with an animal *is* that animal," (2001: 83). However, the animalists' answer to (a) cannot be taken without some sort of qualification. This is because some animalists do not think that personhood is an integral part of animalism. For example, Olson characterizes animalism as a view, according to which, each of us is numerically identical with an animal. That is, there is a certain organism you and it are one and the same (2007: 24). In this view, persons are just material objects (see further Olson 2007; 1997). Moreover, Olson claims that animalism is not an account of what it is to be a person. Consequently, animalism implies no answer to the personhood question (*Ibid.*; also see Merricks 2001: Ch. 5). Olson also claims that psychological continuity is entirely irrelevant for one's survival (1997: Chs. 4 and 5). In light of this, a proponent of animalism may respond to (b) by saying that as long as a human animal is around (i.e., alive), its undergoing a radical change in his/her psychological state would not in any way affect its identity. This is because, in Olson's view, the identity of a human animal is rooted in its capacity to coordinate and regulate its metabolic and other vital functions. (1997: 133). That is, "any organism, persists just in case its capacity to direct those vital functions that keep it biologically alive is not disrupted," (*Ibid.*). From such considerations Olson draws the following general principle:

> If *x* is an animal at *t* and *y* exists at *t**, *x* = *y* if and only if the vital functions that *y* has at *t** are causally continuous in the appropriate way with those that *x* has at *t* (*Ibid.*).

So given the animalist view, the mental aspect of a human person is entirely inconsequential for a person's identity. Of course, for reasons we shall in chapter eight, it is perfectly plausible to say that personal identity is not rooted in one's psychology. But to recognize this, one does not have to embrace Olson's animalism. In fact, for reasons we shall see in chapter six and seven, I reject any attempt to identify a human person with the human animal with which it coincides.

The constitution view theorists' answer to (a) is negative. In this case, for example, Baker argues that human persons are constituted by bodies. But in Baker's view, constitution is not identity (also cf. Doepke 1996: Ch. 7; Johnston 1992: 89-105). That means that a person is distinct from his/her body. What is it that distinguishes a person from her body? In Baker's view, a person is distinguished from his/her body by having a capacity for a first-person perspective essentially (2000: 59). In the constitution view (as characterized by Baker), it seems reasonable to say that if a person were to lose his/her first person capacity, then he/she would no longer be a person. In that case, Baker might respond to (b) by saying that regardless of whatever psychological changes a person undergoes, as long as the first person capacity remains intact, a person remains identical with itself. That said, however, Baker's view seems to be a bit puzzling. This is because, she seems to be equating personhood (i.e., what it means to be a person) with a dispositional property, in this case, a capacity for first-person perspective. For reasons we shall see in chapter eight, identifying a person with a dispositional property of any kind in a way that Baker does is deeply mistaken.

On the other hand, a substance theorist, who takes a person as a substantial non-physical entity, responds to (a) negatively. However, that does not mean that a substance theorist necessarily denies that there is an intimate relation that holds between a person and her animal body. Since for a substance theorist, a person's capacities do not make up a person's identity, her answer to (b) will also be negative. That means that despite experiencing radical change of personality, a person still maintains its identity through it. Along these lines, I will explore my own solutions to (a) and (b), in chapters six, seven and eight. That said, my subsequent discussion of the neo-Lockeans' answer to (a) will be based on Noonan's *Personal Identity* (2003).

Unlike the three views I briefly discussed above, the neo-Lockeans attempt to answer (a) by appealing to various sorts of puzzle cases. In this case, the neo-Lockeans begin their discussion by criticizing the *Bodily Criterion* for personal identity (Noonan 2003: 2). This is the most familiar personal identity criterion that readily occurs to people (Swinburne in Sydney & Swinburne 1984: 3-4). In the normal course of life, we often use the Bodily

Criterion with minimal effort, to identify people at a time and also reidentify them (i.e., over time). So in the Bodily Criterion view, personal identity is constituted by bodily identity or spatio-temporal continuity. That is, if P_2 is a human person at time t_2 and P_1 exists at another time t_1, $P_2 = P_1$ if and only if P_2 at t_2 has the same body as P_1 had at t_1. In saying the same body, however, proponents of the bodily identity are not assuming that the body must contain the same bits of matter over time. Due to metabolic turnover, our bodies do not retain the particles that compose them. However, insofar as the replacement of matter takes place gradually, material bodies can retain their identity through change. Likewise, artefacts can lose and gain different parts without the loss of identity. For example, if I replace the two tires of my car, I still have the same car. Of course, if I replace every part in my car with a different part, then I will no longer have the same car. In light of such considerations, proponents of the bodily criterion of personal identity, do not require matter that constitutes a person at any given time to remain the same over time. So a Bodily Criterion of personal identity allows persistence through gradual change of matter. In this view, personal identity is not different from the identity of other things such as a ship, an oak tree and so on (see further Williams 1973: Ch. 1; Olson 1997; Cf. Wiggins 1967: Part Four).

The neo-Lockeans claim that the Bodily Criterion of personal identity comes apart with certain puzzle cases. One of the puzzle cases used to rule out the Bodily Criterion of personal identity has to do with *the brain*. What is so special about the brain? Of all other parts of the body, the brain is directly responsible in sustaining a person's mental life. We know this because damage to the brain disrupts the functioning of one's normal mental life (see e.g., Popper and Eccles 1977: Chs. 5 and 6). But the same is not true if a person loses one of his/her fingers, kidneys and so on. However, this can only be true as far as it goes. I say this because, if some other vital organs in a person's body are seriously damaged, say of, a liver, lung or heart, the normal functioning of the brain could be disrupted. Contra the neo-Lockeans, the brain's normal functioning is contingent on the normal functioning of other vital organs. Brain is not an organ that can simply have its own life in isolation. So from an empirical standpoint, the brain puzzle case is severely underdescribed (see again chapter four).

Having said that, whenever the neo-Lockeans talk about the brain puzzle case, this often has to do with the brain transfer scenario. In this case, the most popular brain transplant puzzle case is the one that has been suggested by Shoemaker. As Shoemaker puts it:

> One day…a surgeon discovers that an assistant has made a horrible mistake. Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumors, and brain extraction, had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's body and Brown's brain, eventually regains consciousness. Let us call the latter "Brownson." Upon regaining consciousness Brownson exhibits great shock and surprise at the appearance of his body. Then, upon seeing Brown's body, he exclaims incredulously "That's me lying there!" Pointing to himself he says "This isn't my body; the one over there is!" When asked his name he…replies "Brown." He recognizes Brown's wife and family (whom Robinson had never met), and is able to describe in detail events in Brown's life, always describing them as events in his own life. Of Robinson's past life he evidences no knowledge at all. Over a period of time he is observed to display all of the personality traits, mannerisms, interests, likes and dislikes, and so on that had previously characterized Brown, and to act and talk in ways completely alien to the old Robinson (1963: 23-24).

In this brain transplant puzzle case, we are told that Mr. Brown and Mr. Robinson had their brains swapped. This is the modern day version of Locke's soul swap thought experiment. We are also told that the brain swap between Mr. Brown and Mr. Robinson resulted in Brownson. Now we are faced with a tricky question. That is, who exactly is Brownson? Is he Mr. Robinson or Mr. Brown? Notice that the resulting person Brownson has Mr. Robinson's body. Also notice that Brownson's brain is transplanted from Brown's head, who has now Robinson's brain. The neo-Lockeans tell us that, to answer the tricky question above, all we need to do is to pay attention to Brownson's character, memories and personality. In doing so, we can see whose life (i.e., Robinson's or Brown's) Brownson's life coincides with. Since in the imagined case Brownson's mental life entirely coincides with that of Brown's, contrary to a physical appearance of Brownson who has Robinson's body, most neo-Lockeans answer the above tricky question by saying that Brownson must be taken as Brown. Neo-Lockeans claim that what the brain transplant puzzle case shows us is that a person goes wherever his brain goes. That is why, even if Brown's body is left behind, his mental life (psychology) is fully manifested in Brownson. The bottom line here is that, given the brain transplant thought experiment, the Bodily Criterion of personal identity is

undermined. That is, if P_2 is a human person at time t_2 and P_1 exists at another time t_1, $P_2 = P_1$ if and only if P_2 has the same brain as P_1 at t_1. Noonan calls this: *the Brain Criterion* of personal identity (2003: 4).[98]

However, the neo-Lockeans claim that the Brain Criterion of personal identity faces its own problems. To see this, they ask us to entertain another scenario that has to do with removing a part of one's brain without killing a person. This scenario is entertained based on certain clinical cases that deal with severing the fibres that connect the two hemispheres of the brain to treat epileptic patients (see Sperry 1968; Nagel 1971/1975). If one day it becomes possible to remove a whole hemisphere without killing a person, then the Brain Criterion of personal identity fails to hold. In light of such considerations, the neo-Lockeans argue that personal identity is possible without brain identity, since one can survive with only a part of the brain. This proposal does not seem to be far-fetched after all. For example, there is a medical condition known as Rasmussen's syndrome—a condition that damages one side of the brain. If one becomes a victim of this syndrome, then the solution is to do a hemispherectomy, i.e., the removal of half of the brain. For example, one such incident entitled, "Meet the girl with half a brain", had recently been reported by NBCNews.com—an American News media outlet. This is the case where a nine year old girl had her entire right hemisphere removed due to Rasmussen's syndrome (http://today.msnbc.msn.com/id/36032653/ns/today-today_health/t/meet-girl-half-brain/accessed 12/30/2012).

In this case, some philosophers argue that what we need for personal identity is not the whole brain, but part of it. Noonan calls this view, *the Physical Criterion* of personal identity

---

[98] Wiggins modifies Brown/Brownson cases by saying that rather than transplanting Brown's brain in one person, we can transplant each of the hemispheres of Brown into two distinct persons. In that case, Brown cannot be identical to two different people because of the transitivity relation (2001: 209). But Wiggins also realizes that the transitivity relation will not be decisive due to Parfit's new solution (1971). Parfit's argument is that each of the persons that shares one of the hemispheres of Brown stand in relation R to him. From this Parfit draws a conclusion that identity is not what matters in survival. For critical discussion on brain transplant puzzle, see also van Inwagen (1990: Ch. 15).

(2003: 6; see also Wiggins 1967; Parfit 1984: Ch. 10). On this view, what is necessary for personal identity is not the identity of the whole of the brain, but the identity of enough of the brain to be the brain of a living person (Noonan 2003: 6). That is, if P_2 is a human person at time t_2 and P_1 exists at another time t_1, $P_2 = P_1$ if and only if P_2 at t_2 has enough of the brain of P_1 at t_1. In explaining why the Brain Criterion of personal identity must be rejected, Noonan states:

> For in such a case there will be personal identity without brain identity, the survivor only having *part* of the brain of the original person. Admittedly in this case we have the rest of the body to hang on to, so we could appeal to the original Bodily Criterion of personal identity to justify our judgement. But an obvious extension of the case shows that this manoeuvre gets us nowhere. Let us suppose that half of a man's brain is destroyed and then the remaining half transplanted into another body with consequent transference of memories, personality and character traits. Here we can neither appeal to the original Bodily Criterion of personal identity nor to the Brain Criterion to justify the judgment that the surviving person is the brain hemisphere donor. Yet it seems quite clear that if we accept that Brownson is Brown in the original Brown/Brownson case we cannot deny that in this case also the survivor is the original brain hemisphere donor. For if we accept that a person goes where his brain goes it cannot make any difference if his brain in fact consists of only one brain hemisphere combining the functions usually divided between two (2003: 6).

Having ruled out both the Bodily Criterion and the Brain Criterion, the neo-Lockeans claim that the Psychological Criterion of personal identity view is the best alternative (see again §§ 5.1-5.1.3). However, serious objections have been raised against the psychological continuity criterion of personal identity. Here a good case in point is Williams's (1956-72) famous Reduplication Argument. Williams asks us to imagine a person he calls Charles, who after having experienced a radical change of character, one day woke up claiming to be Guy Fawkes. Charles remembers everything Fawkes did. Charles' memory claims can be checked as known to historians. In such a scenario, Williams says that it would be tempting to say that Fawkes has come to life in Charles' body. But Williams claims that this is not what we should conclude from the imagined case. This is because, as Williams further points out, what happened to Charles could also happen to his brother Robert. In this case, Charles and Robert turn out to be two candidates claiming to be Fawkes. Since as a matter of the law of logic, two people cannot be identical to one person, neither Charles nor Robert could be Fawkes. This means that in Williams's view, personal identity is an intrinsic relation. That is

"whether a later individual x is identical with an earlier individual y can depend only on facts about x and y and the relations between them: no fact about any individual other than x and y can be relevant as to whether x is y," (Noonan 1993: xvi). Following Wiggins (1980), Noonan calls this principle, "*the Only x and y principle*," (1993: xvi; emphasis original).

In light of the only x and y principle, Williams claims that even if reduplication does not happen, the identity relation will not hold between, for example, Charles and Fawkes. The bottom line for Williams is that, "the only case in which identity and exact similarity could be distinguished…is that of the body—'same body' and 'exactly similar body' really do mark a difference…the omission of the body takes away all content from the idea of personal identity," (1973: 10). Although as Noonan earlier argued that the neo-Lockeans seemed to have an upper hand over defenders of the Physical Criterion, Williams's reduplication objection shows why that is not true. In light of Williams's objection, the advantage of the Psychological Criterion over that of the bodily criterion remains to be far from obvious. In light of the Williams's reduplication objection, the neo-Lockeans are forced to further modify their psychological continuity criterion (see e.g., Nozick 1981; Noonan 1993). The question remains: what form will their modification take?

Of course, the first thing the neo-Lockeans should do to save their Psychological Criterion is to attack Williams's 'the only x and y principle'. By attacking this principle, however, the neo-Lockeans will inevitably compromise the importance of a numerical identity for personal identity over time. Here the issue comes down to what Noonan describes as the 'Determinacy Thesis,' (2003: 17). We can understand the 'determinacy thesis' to stand for what I called earlier the Butler-Reid Assumption which has to do with the determinateness of personal identity (see again § 5.1.1). One of the most popular ways the neo-Lockeans attempt to get rid of the determinateness of personal identity is by appealing, *inter alia*, to the fission thought experiment. In this regard, Parfit's discussion with Lewis is a good representative.

1. *Parfit and Lewis on Fission*
In his 1971 essay, Parfit sets out to challenge two beliefs: the nature of personal identity

(B1) and the importance of personal identity (B2). For Parfit, self-interest and emotions constitute the primary source of (B1). That is, we are driven to committing ourselves to (B1) simply because (i) we take an interest in our own numerical identity; and (ii) ''when we find ourselves in problem case, we do feel that the question "would it be me?" must have an answer'' (1971: 4; see also 1984: Part III). The most important reason for Parfit's rejection of (B1) and (B2), has to do with the fact that they both entail a numerical identity (as understood from a common-sense point of view). Since Parfit denies a numerical identity as a ground for personal identity/survival, he revises (B1) and (B2). Here, Parfit's revisionism calls for complete deconstructionism/reductionism of the notion of a numerical identity embedded in both (B1) and (B2). The upshot of Parfit's reductionism on personal identity can be summed up as follows:

A. Personal identity consists in psychological continuity/connectedness as opposed to a numerical identity.
B. One's survival must not be made contingent on one's numerical identity, i.e., one's survival must not imply a numerical identity at all.
C. What matters in survival are relations of degree, i.e. *R*.
D. The relation *R* assumed in (a)-(c) must not presuppose a numerical identity.
E. Given (a)-(d), as Parfit sees it, it follows that the revisionism of (B1) and (B2) is justified.

Parfit's E (hereafter *P-E*) has been and still is the subject of a lively debate among analytic philosophers. In this regard, both Lewis and Perry were the early responders to Parfit's 1971 paper (see Rorty (ed.), 1976: Chs. I, III and IV). But Parfit points out that there isn't a substantial difference between his position and that of Perry's. So here I focus on the exchange that took place between Parfit and Lewis. In this regard, Lewis' objection goes right to the heart of P-*E*. Lewis like Parfit holds that, what matters in survival is mental continuity and connectedness. Lewis claims that his present experiences, thoughts, beliefs, desires, and traits of character should have appropriate future successors. His total present mental state should be but one momentary stage in a continuing succession of mental states. These successive stages should be interconnected in two ways: First, by bond of similarity….Second, by bonds of lawful causal dependence (1976: 17). That said, however, Lewis also claims that there is another 'compelling commonsense' answer to the question,

what matters in survival? According to Lewis, the answer is identity, that is, identity between myself, existing now, and myself, still existing in the future (*Ibid.* 18).

Lewis urges that since we have two competing answers to the survival question, we have to choose one. If we had to choose between the two competing answers, for Lewis, the common sense view should be chosen. But Lewis thinks that there is no pressing need for us to choose between the two answers, since the two answers are compatible and both are right. That is the claim Lewis wants to defend (*Ibid.*). This is precisely where Lewis' initial agreement with Parfit ends. Since unlike Lewis, Parfit does not subscribe himself to the claim that the two answers are compatible and both are right. As can be expected, given (*P-E*), a Lewisian solution is not available for Parfit. So as we shall see, Parfit insists that a choice has to be made between the two answers.

Lewis endorses (A) in (*P-E*) above, while expressing his dissatisfaction with (B-C) and eventually with *E*. But Parfit claims that Lewis' response to (*P-E*) is far from convincing. To show why this is the case, Parfit begins by summing up the fission case. As he states it:

> This case [fission] seems to involve three people: the original person, and two resulting people. If we decide that the resulting people are indeed, as they seem to be, two people, we cannot claim that each of them *is* the original person. But we may conclude that 'the relation of the original person to each of the resulting people…contains…all that matters…in any ordinary case of survival.' If this is so—if this relation does contain all that matters, but is not identity—then what matters cannot be identity (1976: 91).

Parfit's main claim here is that the post-fission people cannot be identical to the pre-fission person, since one person cannot be identical to two distinct persons. So for Parfit, the pre-fission person's relation to the post-fission people can only be a relation other than identity. Since for Parfit, a numerical identity is off the table, the only relation that can be invoked here is psychological continuity. As Parfit sees it, Lewis' initial agreement with him on (A) above did not go far enough. This is because, unlike Parfit, Lewis still maintains that (P-*E*) conflicts with our basic commonsense belief about survival. That is, one's future survival requires nothing short of a numerical identity. So the pre-fission person's future survival must consist of both the common sense view of identity as well mental continuity. Such

considerations led Lewis to redescribe the fission case, which Parfit finds it to be seriously problematic.

For Lewis, in the pre-fission state, there are two partially overlapping continuant persons who share their initial stage. Lewis calls the initial stage 'S' and designates the latter two stages as 'S_1' and 'S_2'. Lewis also designates the two continuant persons as 'C_1' and 'C_2'. So 'C_1' consists of S and S_1, and 'C_2' consists of S and S_2, (1976: 25). Furthermore, Lewis makes a distinction between what he calls the R-relation and the I-relation, which he takes to be co-extensive. The former relation being the mental continuity and connectedness among person stages whereas the latter relation holds between the several stages of a single continuant person (*Ibid*. 20-21).

Responding to Lewis' redescription of the fission case, Parfit claims that if we are discussing this case before the division, we can ask: ''Is C_1's present stage R-related to the future stage of C_2?'' (1976: 92). Parfit realizes that such a question might come across as irrelevant, since Lewis claims that when we are discussing people before a certain time, we should count them by the relation of identity-up-to-that–time. But Parfit claims that even if this way of talking about the two post-fission people as *one* could be tolerated, prior to the fission, predictions need disambiguating (*Ibid*). Parfit claims that if we can guess now that it is C_1 as opposed to C_2 who will win the next State Lottery, we can then ask, "Is C_1's present stage R-related to the future stage of C_2?". According to Parfit, the answer to this question must be 'Yes'. This is because given Lewis' own view, C_1's present stage is the stage S; C_2's future stage is S_2 which presupposes that S is R-related to S_2 (*Ibid*. 92). In light of this, Parfit thinks that Lewis' main thesis (i.e., what matters in survival is both identity and R-relation) turns out to be unsustainable. Notice that for Lewis, the identity and the R-relations coincide. But for Parfit, such a coincidence cannot be shown to hold. As Parfit states:

> Can it be true that, as he [Lewis] claims, both that 'what matters to survival' is identity, *and* that what matters to survival is the R-relation? We have just seen that C_1's present stage stands to C_2's future stage in the R-relation. On the thesis, this relation is what matters. But if C_1 now stands in the relation that matters to someone *else* in the future, how can this relation be identity? (*Ibid*.)

Parfit's remark here sums up the gist of his objection against Lewis' attempt to converge identity and mental continuity in one's future survival. What precisely is Parfit's objection? Parfit thinks that Lewis misapplied the original puzzle case. In this case, Sider claims that Lewis mistakenly tried to resolve Parfit's puzzle by claiming that since the I-relation and the R-relation (psychological continuity) are one and the same relation, both can be what matters. But Parfit's original puzzle involved identity, not the I-relation. Hence, Sider claims that Parfit is right in questioning Lewis's claim that the I-relation sufficiently captures the heart of the 'commonsense platitude' (as Lewis calls it) that identity is what matters (Sider in Sally Haslanger and Roxanne Marie Kurtz, 2006: 92-93ff). In his postscript (1983), Lewis responded to Parfit's objection as follows:

> Derek Parfit rejects my attempt to square his views (which are mine as well) with common sense. He objects that before I bring off the reconciliation, I must first misrepresent our commonsensical desire to survive. Consider a fission case as shown. I say there are two continuant persons all along, sharing their initial segments. One of them, C_1, dies soon after the fission. The other, C_2, lives on for many years. Let S be a shared stage at time t_0, before the fission but after it is known that fission will occur. The thought to be found in S is a desire for survival, of the most commonsensical and unphilosophical kind possible. Since S is a shared stage, this desire is a shared desire. Certainly C_2 has the survival he desired, and likewise has what we think matters: mental continuity and connectedness (the R-relation) between S and much later stages such as S_2. But how about C_1? (1983:73).

Regarding C_1 above Lewis says:

> If common sense is right that what matters in survival is identity . . . , then you have what matters in survival if and only if your present stage is I-related to future stages…. If that is right, then C_1 has what he commonsensically desired. For C_1's stage S at time t_0 is indeed I-related to stages far in the future such as S_2. These stages are I-related via the person C_2 (1983:73).

But Lewis' ingenious solution to remedy C_1's survival through C_2's I-related stages raises more questions than it answers. Parfit rightly claims that insofar as C_1's present stage and the future stage S_2 are stages of the same person, it follows that they are both stages of C_2. In light of this, Parfit asks: isn't this the wrong person? (1976: 93). If Parfit's objection here stands its ground, then it seems to me that Lewis' solution all the more looks unconvincing, to say the least. This is because it is not clear how C_1's desire for survival can be said to have

been fulfilled simply by assuming that C_2 is a sufficient condition for it (C_1's desire for survival).

Lewis openly admits this problem in his response to Parfit. For example, Lewis claims that if C_1 had the commonsensical desire that he himself (but not any other person) survives into the future, then C_1's desire is not satisfied. Yet despite that, Lewis still thinks that C_1 could not have had exactly that desire, i.e. the desire for survival. But why does Lewis think that is the case? Lewis describes the reason for this as follows:

> And there is a limit to how commonsensical one's desires can possibly be under the peculiar circumstance of stage-sharing. The shared stage S does the thinking for both of the continuants to which it belongs. Any thought it has must be shared. It cannot desire one thing on behalf of C_1 and another thing on behalf of C_2. If it has an urgent, self-interested desire for survival on the part of C_1, that very thought must also be an urgent, self-interested (and not merely benevolent) desire for survival on the part of C_2. It is not possible that one thought should be both. So it is not possible for S to have such a desire on behalf of C_1. So it is not possible for C_1 at t_0 to have the straightforward commonsensical desire that he himself survives. If C_1 and C_2 share the most commonsensical kind of desire to survive that is available to them under the circumstances, it must be a plural desire: let *us* survive (*Ibid*).

As I see it, Lewis has not yet given us a convincing reason why C_1's desire for survival cannot stand on its own right independent of that of C_2's desire for survival. All that Lewis has tried to do so far is to force us to agree with him that C_2's desire for survival is all that we need to make sense of C_1's desire for survival. But this begs a question in that the reason why C_1's desire for survival is merged with that of C_2's desire for survival is because the latter has already been thought to be sufficient for the former's desire for survival. But this is no more convincing than trying to calm down a hungry man by telling him/her that if someone eats food on behalf of him/her that his/her hunger will be satisfied.

But there is a deeper problem with Lewis' attempt to spell out a commonsense notion of identity in terms of stage-sharing. As we recall, the commonsense notion of identity is understood in relation to persons as opposed to stages. In fact, this is clearly delineated in Parfit's own discussion. But Lewis's move to link identity with stage-sharing does not advance his case. In talking about this problem, for example, Sider asks: how exactly does Lewis understand the commonsense platitude? Sider proposes one possible answer which I

think clearly captures Lewis's point. As Sider puts it, "A person stage matters to my present stage if and only if it bears the I-relation to my present stage," (Sider in Haslanger and Kurtz, 2006: 93). Sider claims that here Lewis's point concerns what matters to person stages. But Sider claims that our commonsense notion of 'what matters', is understood in relation to what matters to persons. In light of this, Sider claims that Lewis can only vindicate the commonsense platitude that identity is what matters, only if his version of that platitude must concern what matters to persons as opposed to person stages. Since for Lewis persons are not person stages, his emphasis on person stages does not address the present topic (*Ibid*. 93; see further Lowe 1998: Ch. 5; Oderberg 1993: Chs. 2-5).

2. *The Stalemate*

At this point, the hitherto discussion might well bring to our mind Rorty's observation. As we recall, Rorty told us that, "disagreements about the criteria for personal identity have been persistently unresolved, the battle lines repetitively drawn over the same terrain, along familiar terrain, along familiar geographical strongholds," (1976: 1). This is partly because the contemporary debates on personal identity are based on extremely underdescribed puzzle cases on the one hand, and mutually exclusive claims made with respect to what counts as the criterion of identity on the other hand. Moreover, thought experiments are proposed with no clear guidelines, for example, such as the ones I had suggested in chapter four, namely the *Role Condition*, the *Constraint Condition* and the *Priority Condition*. So in personal identity discussions, philosophers can come up with any thought experiment they like to defend a particular view of the notion of personal identity. This makes it hard to have a common ground in discussions besetting personal identity. This is one reason why, as we saw above, it is exceedingly difficult to settle the debate between the neo-Lockeans and their opponents. Moreover, the contemporary personal identity debate also tends to be far removed from lived experience. As Schechtman remarks:

> The personal identity problem has enjoyed a revival among analytic philosophers over the last three decades. Since questions of personal identity are of fundamental interest outside philosophy, there is some reason to hope that in this area philosophy will do what it is popularly thought to do—apply rigorous standards of argument and investigation to basic problems of human existence. A glance at the contemporary literature on personal identity, however, quickly disappoints these expectations. Instead of questions of self-knowledge, self-expression, and

authenticity, we find discussions of the necessary and sufficient connections between entities called individual "person time-slices" which allow us to say they are slices of the same person. These creatures inhabiting philosophical theories of identity seem to have little to do with persons as we know them, and the concerns about identity these theorists address seem far removed from the compelling identity issues familiar to us from lived experience, psychology, and literature. (1996: ix).

Schechtman's observations in the above passage coincide with my own observations that I have been making throughout this thesis. Schechtman's observations are also in line with Rorty's earlier observations. That said, however, Rorty's own observations concerning the state of the contemporary personal identity debate are not free from their own shortcomings. For example, of the four questions, namely class differentiation, individual differentiation, individual reidentification and individual identification, Rorty claims that, if philosophers distinguish which ones are centrally interesting, then progress can be made on the personal identity debate. I disagree. This is because, however important these questions may be in their own right, none of them are central to the metaphysics of personal identity. Rather the central question that we need to focus on, to make progress on the personal identity debate, is what I called in chapter one, the *Aristotelian Question*. Rorty's four questions are subservient to the *Aristotelian Question*, since they all presuppose it. So unless the *Aristotelian Question* is taken seriously, it is difficult to see how we can engage in a substantive discussion concerning personal identity. In this thesis, I am very much interested in shifting our focus from such a repetitive trend of debating about the nature of personal identity based on science fiction thought experiments to taking ontology and metaphysics seriously in thinking about the nature of the self/person.

C. *Outstanding Issue*

Besides the complications science fiction cases add to the existing lack of agreements on the criterion of personal identity, the neo-Lockeans notion of a person is no less obscure than that of Locke's. The leading neo-Lockeans (Lewis, Perry, Noonan, Shoemaker *et al.*,) are all materialists. But the compatibility between their materialism and their neo-Lockean conception of a person is by no means obvious. For example, if as neo-Lockeans argue that in the case of brain transplant "a person goes wherever his/her brain goes," then it follows

that (i) a person is distinct from his/her brain; (ii) a person is also distinct from his animal body; and (iii) a person is a thinker who is distinct from a set of properties it bears. But (i)-(iii) together support a dualist view of a person as opposed to a materialist view of a person. I will defend this claim in chapter six. But in this regard, there is ongoing debate between animalists and the neo-Lockeans. The former arguing for the claim that there is only one thinker, that is, a human animal (see for details Olson 1997; 2007). On the other hand, the latter arguing for the claim that only a person is a thinker (see e.g., Shoemaker 1984: Part Two). But the neo-Lockeans reject animalism, claiming that persons are not human beings owing to their different persistence conditions. However, recently Noonan argued that the neo-Lockeans must allow for the possibility of multiple thinkers (see e.g., Noonan in Gasser and Stefan 2012: 89-95). Noonan's point seems to be that the animalists as well as the neo-Lockeans can have it their own way by allowing the possibility of too many thinkers. But it remains unclear how Noonan's proposal here says anything helpful by way of settling the outstanding issue of what precisely is the ontology of a person, given the neo-Lockeans commitment to materialism.

D. *Summary*

In this chapter, I attempted to show that the neo-Lockeans emphasis on the criterion of personal identity and the puzzle cases sidelined the Aristotelian question. I also attempted to show why the debate between the two dominant rival views, namely the Psychological Criterion View and the Physical Criterion View is difficult to resolve. I then suggested that the way out of this impasse is to shift our focus to taking metaphysics seriously in thinking about the nature of a self/person. I now turn to chapter six, where I will attempt to do that.

Chapter Six

Jackson's Location Problem (Serious Metaphysics) and Argument from the Self/Person

In chapter five, I argued that the contemporary personal identity theorists' unduly emphasis on the methodology of thought experiments on the one hand and the criterion of personal identity on the other, for the most part is responsible for the present stalemate in contemporary personal identity debates. I also argued that what I called the *Aristotelian Question*, is neglected in the contemporary personal identity discussions. In light of such considerations, in chapter five, I tried to flag out why it is important to take the ontology of self/person seriously. In the present chapter, my goal is to give some reasons why we need to take the ontology of the self/person seriously. That said, in this chapter, I will advance my discussion of the self/person within the context of philosophy of mind. I will do this by taking two views, one from the dualist camp and the other from the strict physicalist camp. My discussion in this chapter is also intended to show us how it is possible to carry on a sensible discussion of the self/person without entangling ourselves with puzzle cases. Ultimately, my discussion in this chapter paves the way for my own view of the self/person which will be introduced in chapter seven.

To what is the self or person identical? Are persons identifiable with an organic body? Or as Lowe argues do persons constitute a distinct sort or kind of entity and are not to be identified with the biological entities in which they are embodied? (1989:2).[99] Such questions in the contemporary literature in philosophy of mind are treated in different ways. As I see it, at least three ways stand out: (1) Keep the self or person distinct from one's body (e.g. Lowe, 2008: Part I).[100] (2) Avoid the self or person and focus only on consciousness (e.g. Churchland, 1988 and Kim, 2006). (3) Eliminate the self or person or locate it within a naturalist ontology which entails a strong version of physicalism (e.g.

---

[99] See further, Lowe, (2008).
[100] Also see similar argument in Lowe (1996: Chs. 1, 2 and 7; 1989a:2; 2000: Chs. 2 and 10).

Jackson 1998: ch. 1).[101] Let us call (1) the thesis about the self (TS), (2) the thesis about consciousness (TC) and (3) the thesis about serious metaphysics[102] (SM) respectively. TS and SM will be the main focus of our discussion. In this chapter, I will not engage directly with TC except briefly in relation to some objections I consider against TS.

Before we move forward, here are some important distinctions and definitions of key terms used throughout this chapter. According to Lowe, a self or person is a being that is not identifiable with his or her organic body, nor with any part of it, such as the brain (Lowe, 2008: 19ff). In short, a 'self' or 'person' is distinct from the physical body that embodies it. The precise development of such a characterization of the self or person will become clearer as we go along in our discussion. In light of this, my main purpose in this chapter is to approach the question of 'the self' or 'person' as stated above in (1). Moreover, to meet my purpose, throughout this chapter, I assume (unless otherwise indicated) a version of dualism Lowe calls *Non-Cartesian substance dualism* (hereafter, NCSD) which further expands and develops (1). As Lowe puts it, according to NCSD, "it is I, and not my body nor any part of it, who am the bearer of mental properties…." (*Ibid*: 95).[103]

On the other hand, by philosophical naturalism, following Craig and Moreland, I refer to the thesis that the entities that exist are the ones that are in the spatio-temporal universe as studied by physical sciences (Craig and Moreland, eds., 2000).[104] In light of this, by a strong version of physicalism, I refer to a thesis according to which the physical nature of our world exhausts all that exists (e.g. states of affairs, all individuals, minds, semantic content, properties, laws, etc.). Put in this way, the philosophical naturalists I have in mind are those who advance the strong version of physicalism.[105] I now turn to Lowe's theory of the self.

_____

[101] For a crucial distinction on the naturalist ontology, see, Moreland (2008: Ch.1).

[102] *'Serious Metaphysics,'* is the term Jackson himself uses. I will explain his view in § 6.2.

[103] Though NCSD shares some similarities with Cartesian substance dualism (CSD), it differs from it in a significant way as I will show at some point in the course of our discussion. Given NCSD, in this chapter my main aim is to argue that strong physicalism of Jackson's version fails to account for the nature of the self.

[104] Also see Papineau (1993). I also use the term 'philosophical naturalism' interchangeably with 'scientific naturalism.'

[105] All philosophical naturalists are not advocates of strong physicalism. For example, John Searle endorses various emergent properties and hence, Searle can be called a 'weak naturalist,' see Moreland

6.1 *Lowe's Theory of 'The Self'*

Lowe argues, in his *Personal Agency*, that "a person or self does not appear to be simply identifiable with his or her organic body nor with any part of it such as the brain—and yet selves seem to be agents, capable of bringing about physical events, such as bodily movements, as causal consequences of certain of their conscious mental states," (2008: 19).[106]

As I see it, central to Lowe's characterization of a person or self, *inter alia*, is the idea that a person or self is not identical with one's physical body nor with any part of it, such as the brain.[107] For example, suppose that X is identical with Y. What does this mean? At the least, it means that X and Y are numerically one and the same thing. In this case, no distinction holds between X and Y. But if we say that X and Y are not identical to each other, then that means that X and Y are not numerically one and the same thing. Hence, a distinction holds between X and Y. Behind this assumption is the familiar Leibniz's law of the *indiscernibility of identicals*, according to which for any X and for any Y, if X and Y are identical, then for any property P, P will be true of X if and only if P is true of Y. The gist of this law is that if X and Y are identical, then X and Y share one and the same properties. What we have is not two entities but one and the same entity (see again chapter one § 1.2.2 A).

One way a proponent of strong physicalism may challenge Lowe's TS is by showing that the self is identical with the physical body which embodies it. If strong physicalists succeed in meeting this demand, then the self and the physical body are numerically one and the same thing. Hence, when we talk about the self and the physical body, we are literally talking about one and the same thing, that is a self just is a physical body and vice versa. But things do not seem to be that easy when we carefully reflect on the kinds of distinctions that we make. For example, as P.F. Strawson remarks, "Each of us distinguishes between himself

_____

(2008: Ch.1); Searle (1992). For strong version of physicalism, see, Jackson (1998: Ch.1) and Papineau (1993: Part I).

[106] For a similar line of argument, see Lowe's other writings as indicated in footnote number 99.

[107] Churchland argues that Artificial Intelligence and Neuroscience play a key role in unfolding the nature of the human brain. So he rejects the notion of 'the self'; see Churchland (1988: Chs. 6-8).

and states of himself on the one hand, and what is not himself or state of himself on the other…." (1959: 87ff). If Strawson is right in his remark, and I believe that he is, then our experience of making distinctions would pose a formidable challenge to the proponent of strong physicalism. This is precisely because for strong physicalists, the idea of the self being distinct from the physical body is irreconcilable with their naturalist ontology, which endorses only entities that exist in the spatio-temporal universe as studied by physical sciences. Since the self is not something that can be established on the basis of an empirical science (for reasons we will see later in this chapter), the strong physicalist will inevitably be forced either to deny that there is such a thing called the self or engage in some kind of reductive analysis of the nature of the self. I will say more on this in § 6.2.

It seems to me then that to vindicate Lowe in his claim that the self is distinct from the physical body that embodies it, we need to be able to point out one thing that is true of the self but that is not true of the physical body and vice versa. If we could do that, then we can show that the self is not identical with the physical body. If I am right here, then the physicalist can utilize Leibniz's law[108] as stated above for his or her benefit only if the self is the same as the physical body or some part of it.

Returning to Lowe's TS, the key question that we need to answer is this: *What is the self?* As Lowe sees it, this question can be understood either (1) as a request for the meaning of the term 'self', or (2) as a request for an account of the nature of the self. Putting aside the details for the present, what we need in order to understand Lowe's notion of self is to approach the question of the self from the standpoint of (2) (Lowe, 1996: 182ff).[109] Taken this way, Lowe claims that the obvious place to start such a discussion is with *first-person reference*. That is, as Lowe puts it, "A person or self…is a being that can have thoughts about itself, of the sort that are appropriately expressed (in English) by sentences containing the first-person pronoun, 'I', as their grammatical subject-sentences such as 'I feel hot' and 'I

---

[108] Both Kim and Churchland conflate Leibniz's two laws on identity as stated above. See, Kim (2006: 35 and 39); Churchland (1988: 29).

[109] Of course, opinions differ on what 'a self' is but getting into that discussion is not my interest here. For present purpose, I am only interested in Lowe's take on the self. For more on this issue consult the reference in chapter one under footnote # 2.

am six feet tall'," (Lowe, 2008: 20-21).[110] In short, for Lowe, a self is a possible object of first-person reference that is capable of referring to itself as 'I', (Lowe, 1996: 5-6 and 183). So, on this view, a necessary condition for selfhood is to have a capacity for self-reference using the first-person pronoun 'I'. But Lowe's point here could easily be misunderstood if one thinks that anything can be a self insofar as it utters a sentence which can express a first-person thought.

For example, we may read on a computer screen a sentence that can express first-person thought such as 'I am shutting down', or 'I am having some difficulty taking your order', etc. But as Lowe claims, from this it does not follow that a computer is a self. A computer, despite displaying sentences on its screen that express first-person thought, is not thereby expressing literally the thought that the computer itself is 'shutting down' or 'having some difficulty in taking someone's order', etc. But strong physicalists, who deny Lowe's TS, still argue that there is no difference between persons and computers. For example, Churchland claims that it is possible to produce and configure a purely physical device which can possess genuine intelligence. According to Churchland, machines like computers insofar as they simulate all of humans' cognitive activities to the last computational detail, can be conferred the status of genuine persons (1988: Ch. 6, see esp., 99 and 119-120).

However, Churchland's remarks here remain unpersuasive. As Alexander Pruss recently argued that persons have objective as opposed to socially defined identity conditions whereas the same can't be said with respect to the so called 'electronic persons' such as computers or robots (2009: 487-500). All things being equal, all that computers can do is to process the input that has been fed into them within the parameter set for them. So, for instance, when a computer displays a sentence on its screen that expresses a first-person thought, a computer is simply executing some sort of order as an output in accordance with the way it has been programmed to do so.[111] But if Lowe's TS is correct, as we shall see, such is not true of persons or selves.

---

[110] For different views on 'I', see, Madell (1981: Ch. 1).
[111] For a similar argument, see, Kim (2006: 142-145), where he discusses the Turing test often used to show that purely physical devices can think like persons. But Kim rightly rejects the Turing proposal on

However, strong physicalists may still insist that complexity is all that we need to get something in the neighborhood of a person. Thus, the argument goes, insofar as computers or purely physical devices reach a sufficient level of complexity, then we can attribute to them a genuine intelligence. But as Searle points out (1984: 31-32), such a thing seems not to be the case. No matter how complex a computer may be, it would still not be the case that a capacity to exercise a mental state (e.g., thinking) can be put on par with a computing machine that runs a program. As Searle rightly remarks, the computer's ability to manipulate a set of rules or syntax does not show in any clear and plausible way that a computer grasps the semantics, that is, the meaning or what a set of rules represent.[112] If Searle is right here, then it follows that the capacity to grasp semantics or meaning is exclusively something that only a person or self has. If this is correct, then this will give a further impetus for us to endorse Lowe's TS.

However, here I am not implying that Searle endorses Lowe's TS, but that is beside the point. It seems then highly counterintuitive to think that a computer (by virtue of attaining a sufficient level of complexity) can have a concept of itself, for example, 'shutting down' or 'having difficulty', etc. Hence, it is plausible to conclude as Lowe claims, that computers neither have a capacity for self-thought nor any thought at all (1996: 183). Thus, the argument that purely physical devices like computers can be granted the status of a self is far from conclusive.

Since for a proponent of NCSD the self is distinct from the physical body or any part of it, it seems plausible to say that the capacity for first-person thought can only be had by the self, not by the physical body or any part of it. But notice that having a capacity for first-person thought per se is not sufficient to substantiate Lowe's TS. In other words, there is more to a

_____

grounds that the input-output equivalence does not indicate the presence of mentality in physical devices. See Lowe's discussion (2000: 209-214). See also McGinn's view on this (1999: 175-196).

[112] In fact, Searle's point here can be made without any recourse to thought experiment. However, one need not endorse the 'Chinese Room thought experiment' to get Searle's point. For Searle's full account of the thought experiment, see Searle (1984: 31-32). I also think that the 'frame problem' effectively shows that the machines neither have the knowledge nor the ability to act in a way that fits the relevant circumstances. But human beings have both of these things. For more on this and on 'connectionism and the mind' see, Lowe (2000: 218-229).

self than merely having a capacity for first person thought. In this regard, at least, there are three things worth mentioning. These are self-knowledge, agency and the person as an emergent substance, respectively.

6.1.1 Self-knowledge

Lowe claims that a self has first-person knowledge about itself. For Lowe, what a self knows about itself has to do with knowing that certain thoughts, experiences and actions are its own. That is, as Lowe puts it, "it is a logically necessary condition of selfhood that a self should know that it itself is the unique subject of certain thoughts and experiences and the unique agent of certain actions," (1996: 183). Lowe's point here is that there is a sense in which the self necessarily knows that the self itself, rather than someone else, is the unique subject of certain experiences and thoughts. But as Lowe claims, the question remains: "which thoughts and experiences are they of which the self necessarily knows that it itself is the unique subject?" (*Ibid*: 184). Here, different answers can be given, some of which Lowe raises (*Ibid*: 184). But for present purpose I will not get into that discussion.

Notice however, that the question Lowe raises here seems to presuppose that there is a particular kind of knowledge. If the self has it, then it follows that such knowledge is genuine self-knowledge. In light of this, Lowe asks us to distinguish between *de re* and *de dicto* knowledge. Here the terms *de re* and *de dicto* are both modal notions. That is, in the words of Loux, "whereas the ascription of a *de dicto* modality is the ascription of the property of necessary truth/falsehood, possible truth/falsehood, or contingent truth/falsehood to a proposition taken as a whole, the ascription of a *de re* specifies the modal status of a thing's exemplification of some attribute," (1998: 184). Yet the distinction drawn here between the *de re* and *de dicto* modalities is not welcomed by all. In fact, there are philosophers who reject the *de re* modality in favor of the *de dicto* modality (e.g. see Quine, 1963: 148; 1960: 199). But for present purpose, I will not belabor Quine's objections against the *de re* modality even though Quine's own objections are far from conclusive (e.g. see Davidson ed., Plantinga, 2003: Ch.1).

According to Lowe, the *de dicto* knowledge does not show that the self has a genuine self-knowledge. The reason for this is that the knowledge claims that we make in the *de dicto* sense arise purely from our understanding of the fact that thoughts and experiences are individuated by their subjects. Thus, Lowe claims that the *de dicto* knowledge claims barely qualify as expressions of self-knowledge. So this leaves us with *de re* knowledge, according to which one's knowledge claim has to do with a thing or an object. For example, I necessarily know of the thought or experience of presently writing this thesis, that it is my own, that I (and no one else) am the subject of this thought or experience. Thus, as Lowe puts it, "when I have a conscious first person-thought—such as the thought that I feel hot— I regard myself as being the subject of this thought, both in the sense of being the thing having the thought and in the sense of being the thing that the thought is about," (2008: 21). But Lowe is not claiming here that one also necessarily knows the other person's present conscious experiences or thoughts, of which the other person is their only subject. In light of such asymmetry between knowledge of self and knowledge of others, for Lowe only the *de re* knowledge is a genuine instance of self-knowledge (1996: 185; cf. Coliva, ed., 2012).[113] Here Lowe's remarks remind us of our discussion of Descartes's *Cogito*, which as I argued is rooted in *de re* knowledge (see chapter one § 1.1.1.1).

So assuming that the self has self-knowledge, the question remains: what then can we say about the body that embodies the self, that is, can we say that one's body or one's brain is also the subject of the thought or experience? If Lowe's TS is correct, then this question must be answered in the negative. This is because only the self is the subject of thoughts. This is both in the sense of the self being the thing having the thought and also in the sense of the self being the thing that the thought is about (2008: 21). It then seems that we can deny that the body or any part of it (e.g, the brain) is the subject of the thought or experience. For instance, when I think of doing something, I do not understand my 'thinking of doing something' in terms of my body or any part of it 'doing the thinking' I am engaged in. Similarly, if I make the statement, 'I hate so and so', then it would not make any sense to

_____

[113] Lowe extends this discussion beyond what I have covered here, see (*Ibid*. 185-203).

understand it in terms of my organic brain (despite using it to think) or my hands or my eyes, etc., as the subject of the thought of hating. So, in this case, it must be the self that is the subject of the thought of hating. Since the 'thought of hating' itself is a non-physical mental state, it is also distinct from the physical/brain states (e.g., neuronal activity such as C-fiber stimulation). Put differently, mental states are to the self as the physical states are to the physical body or any part of it. I will say more on this as we proceed in our discussion. Given what has been said so far, it seems plausible then to assume that a property of self-knowledge necessarily belongs to selfhood.

Hence, from this it follows that the self cannot be made identical with the body that embodies it. Thus, self-knowledge is one of the qualities that we attribute only to the self as opposed to the physical body or any part of it. In light of this, we can say that we have here a strong ground to justify the claim which I have been advancing from the outset. That is, in order to show whether a distinction holds between the self on the one hand, and the body or any of its parts on the other, we must point out one thing that is true of the self, but not true of the physical body and vice versa. As it seems then, self-knowledge is one of those things that is true of only the self, which is not true of the body nor any part of it, such as the brain. Hence, we can say that in light of this, Lowe's TS is tenable.

6.1.2 Agency

Lowe claims that the self is by its very nature an agent, that is, the self is naturally capable of performing intentional actions, some of which bring about physical results (2008: 23ff). On Lowe's view, an action *A* is said to be an intentional action, just in case an agent performs it in a certain way knowing that an agent is so acting. It seems to me that here, part of what constitutes being an agent, is the ability to perform intentional actions. But this is not enough. Lowe claims that for an action to count as intentional, an agent must also know that he/she is performing a particular action in a particular way. But here, I am not saying that for Lowe all there is to be an agent is the ability to perform intentional actions. However, performing intentional actions is one of the crucial aspects of the self's nature. We can then ask: can we draw a clear demarcation between intentional actions vs.

194

unintentional actions? Different philosophers have different opinions on this. In this chapter, I will not pursue this issue (see Lowe 2000: Ch. 9).

As we recall, for Lowe, the self is necessarily capable of self reference (e.g., via the use of the first-person pronoun 'I'). In light of this, Lowe claims that the self necessarily possesses agency, that is, the ability to do intentional actions. Lowe sees self-reference as a species of intentional action (2008: 24). Thus, as Lowe puts it, "to refer to oneself as 'I', whether in speech or merely in thought, is to perform a kind of intentional act," (*Ibid*: 24). The point here is that to refer to oneself as 'I' presupposes one's engagement in some sort of intentional act. As Lowe sees it, if done only in thought, such an act may not have physical results. But if done in speech, such an act will bring about physical results (*Ibid*: 24). Furthermore, for Lowe, only the self (capable of performing intentional actions) can have a concept of causation. In light of this, Lowe claims that possessing such a concept of causation is a necessary condition of self-reference. Thus, on Lowe's view, possessing a concept of causation is linked with selfhood itself. Since an intentional action itself is a causal concept, when an agent performs an intentional action, Lowe thinks that an agent thereby also causes some sort of event to take place. This in turn will have a physical result, unless otherwise done merely in thought.

So, for Lowe, the key notion here is that selfhood requires agency. By 'requires agency' here I understand Lowe to mean that either the self is the primary source of an intentional act or that it is the self or person that is capable of bringing about intentional acts. If this is correct, then it follows that the physical body or any part of it does not require agency in a sense stated above. For example, suppose that 'I consciously desire a Starbucks coffee. Here at the least, I can say that I am in a Starbucks coffee 'desiring state'. But to be in such a state is not in any clear way to be in a physical state. If this is correct, then to have a desire for a Starbucks coffee, in this case, pertains to a purely mental state. Thus, a mental state does amount, for example, neither to c-fibers firing inside my brain nor to some sort of publicly observable behavior. However, here I do not want to deny that there is a correlation between a mental state and a brain/physical state. For example, if I am a coffee addict,

then, when I drink coffee my headache goes away or I can better focus on doing something, etc. Here, I put aside the issue of whether such correlation is contingent or necessary.

So it seems pretty clear that my desiring a Starbucks coffee (mental state) coupled with drinking coffee will result in some sort of physical experience (e.g., having no headache, etc). However, the question remains: what is it that, in the first place, makes me go to Starbucks to drink coffee? We may say that it is primarily the desire I have (not to mention the belief I also have of the existence of the Starbucks coffee shop itself at a specific location, etc.) to drink the Starbucks coffee. But here I do not take my 'desire' to drink coffee as necessarily playing a decisive role in my being at the Starbucks coffee shop. This is because, despite my desire to drink coffee, I can suppress my desire and fail to act on it.[114] I assume that we do this with a host of other issues as well.

Therefore, as Lowe rightly argues, all things being equal, to perform an intentional action, an agent or the self first must *will* to do so.[115] In this case, I must *will* to go to the Starbucks shop knowing that I am going there to drink coffee. So, in light of the desire I have to drink coffee, what enables me to go to the Starbucks coffee shop is my own *willing* to do so. In other words, being an agent *per se* is not sufficient to bring about an intentional action which has a physical result. An agent also needs to engage in *volition* or in an *act of will*. But for one to engage in an *act of will*, at the least, one should be able to exercise a power to choose otherwise.[116] So, given that the self or person is by its very nature an agent that is capable of performing an intentional action, we can point out human *agency* as one of those things that is true of the self but not true of the physical body or any part of it. Thus, we can say that the self is not identical with the physical body nor any part of it. From this

---

[114] There are philosophers who argue that beliefs and desires cause our actions. See, for example, Kim (2005: Ch.1).

[115] Lowe provides a detailed analysis on agent causation, event causation and volitionist theory, he himself favoring the latter one. See on this, Lowe (2008: 121-212; 2000: Ch.9).

[116] This issue is very controversial because not everyone accepts that humans are free in a libertarian sense, which is the ability to do otherwise. Since my sole purpose here is to deal with how the notion of agency implies Lowe's TS, I do not want to digress from my main issue in pursuit of something else off my main agenda. But I will return to this issue briefly in § 6.4.1. For debate on free will, see Kane, ed. (2002).

then Lowe's TS follows.

6.1.3 *The Self as an Emergent Substance*

Lowe claims that the self or person is a substance. By 'substance', Lowe has in mind a traditional notion of substance as a concrete individual thing or continuant. Here an example of substance would be an individual dog. A substance is said to be continuant just in case it persists through qualitative change. For example, if X's skin color goes from being black to being white via cosmetic surgery, then the very same individual X must endure such a change and remain the same through such a change (see chapter one § 1.2.2.B; see also Lowe 1996: Ch. 1). In light of this, *inter alia*, a substance underlies a change a property undergoes. A lot can be said here, but for present purpose I want to focus only on some aspects of Lowe's discussion on substance. I will pick up this discussion again in chapter seven where I will sketch out my own view.

As Lowe argues, the self is a psychological or mental substance, whereas the physical body is a physical or material substance. Since a substance is a bearer of properties, Lowe rightly claims that a mental substance bears mental properties, whereas a physical substance bears physical properties. But does this mean that these two substances are radically different? Here different answers can be given. For example, as we shall see in §§ 6.2 and 6.3, a strong physicalist only endorses a physical substance at the expense of dispensing with a mental substance. On the other hand, for Descartes, the mental substance (the mind) is radically different from that of the physical substance (the body).[117] Why think this is the case? One reason is that (as we recall from our discussion in chapter one), for Descartes, a mental substance is purely *immaterial* whereas the physical body is purely a *material* substance. Thus, for Descartes, a mental substance bears only mental properties. In other words, the self cannot have a physical property. One of the obvious serious problems here, of course, would be how to make sense of the causal interaction between two such radically diverse substances. In my opinion, this remains a tough obstacle for the plausibility of the Cartesian

---

[117] See, for example, Rene Descartes, ed., by Donald A. Cress, *Discourse on Method and Meditations on First Philosophy* 1998: 63-69; 92-103. See also Lowe (2000: Ch.2).

model. But this is not my concern here. By contrast, according to NCSD, the two substances raised above are not radically different since the self or person is embodied by the physical body, yet distinct from the physical body or any part of it. The question remains: how should we understand the claim that the self or person is an emergent substance?

In the literature, the term 'emergence' is employed in two distinct senses, namely strong emergence and weak emergence. As Chalmers explains, a high-level phenomenon is strongly emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain. Yet, truths concerning that phenomenon are not deducible even in principle from truths in the low-level domain. Strong emergence is often discussed in philosophy. By contrast, a high-level phenomenon is weakly emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain. Yet, truths concerning that phenomenon are unexpected given the principles governing the low-level domain. Weak emergence is often discussed in science (Chalmers in Clayton and Davies 2006: Ch. 11). I will take up this issue in Chapter eight in some detail.

For Lowe, the self is an emergent substance in the strong sense of emergence. In this case, the self or person is ontologically distinct from the physical body that embodies it (2008: 92-118; cf. Hasker 1999).[118] To say that the body embodies the self is to say that the body concretizes it—i.e, makes it concrete, tangible or visible. This is contrasted with the constitution view. In this view, to say that the body constitutes the self/person, is to say that the body is related to the self/person as the statue is related to the piece of clay that constitutes it. In this case, the two objects occupy the exact same region of space yet with their own persistence condition (see again Ch. 1: § 1.1; Baker 2000). By contrast, the embodiment relation is based on what I shall call the 'having relation'. That is, the self/person has the body that allows him/her to move, to sense, to see, to hear and so on. Although particles constitute a person's body, given the embodiment model, the same cannot be said of the self/person itself. In other words, nothing constitutes the self/person (see further Lowe 1989).

---

[118] By 'emergent substance', I understand Lowe to mean 'a new kind of substantial entity.'

In light of this, Lowe claims that the self has a causal power. But such causal power attributed to the self is not reducible to the causal power of the physical body that embodies it. Nor does such a causal power, on the other hand, exist totally independent of the causal power of the physical body. Thus, Lowe sees the causal power of the self as complementing and supplementing that of the physical body, as opposed to being in conflict with it. Unlike Descartes who thinks that the self has no physical properties, Lowe claims that there is a sense in which the self is a 'physical' thing. That is, the self is a thing that possesses physical characteristics such as height—even though it has different identity conditions from those of the body or brain (*Ibid*: 22).

To see Lowe's point more clearly here, we need to understand that for Lowe, the self is a simple substance. We say something is a simple substance just in case it is not composed of parts. But for X to have a part, X itself must be a substance. For example, our physical body (material substance) is composed of a number of parts, namely hands, legs, eyes, etc. So we can say that the eyes or any other part of the physical body taken individually is a 'substantial part' of the material substance (in this case, the physical body) taken as a whole. But since for Lowe, the self is a simple substance, it does not have such parts. In light of this, Lowe distinguishes a substantial part of a thing from a merely spatial part of it (Cf. Heil 2012: Ch. 2). It is the latter sense that Lowe employs for the self. That is, the self has a spatial part, where 'spatial part' refers to some geometrically defined section of it. However, here the word 'section' should not be taken to mean something is literally cut out from the self. Instead Lowe tells us that when we refer to the self's spatial part, we refer to it only in a sense that the self has a region that is defined by certain purely geometrical boundaries (1996: 36). Hence, when we attribute a physical characteristic to the self, such as 'X is five feet tall', all we are saying according to Lowe, is that the self has spatial parts as opposed to substantial parts. Since the self has no substantial parts, it is distinct from the physical body or any part of it.

Moreover, for Lowe, the self is a *subject of experience*. In this context, by 'experience' Lowe means that not only does the self have sensory and perceptual experience but also inner awareness and thoughts (2008: 94; cf. Moreland 2009: 110-111). Since only the self bears

mental properties, it must be distinct from the physical body or any part that embodies it. Of course, the self exemplifies mental states (e.g., pain), when the physical body or part of it is in a certain physical state. But the question remains: does this imply that the self must have a physical body or any part of it in order to have a mental state? Lowe thinks not. That is, as Lowe sees it, "I being the subject of all and only my own mental states, am such that every one of those mental states does depend upon me…hence…neither my brain as a whole nor any part of it can qualify as the subject of all and only my mental states and so be identical with me," (2008: 97). If Lowe is right here, which I believe that he is, then bearing mental properties is something which is true of only the self but not true of the physical body or any part of it. So once again this shows that Lowe's TS is plausible.

In light of what has been said so far above, I believe that Lowe's theory of the self makes a strong case for TS. However, the strong physicalist may employ some ways to object to Lowe's TS. Such an objection could be raised, for example, by suggesting that we can only have a plausible understanding of the nature of the self, if and only if, we either locate it within the framework of physicalism or eliminate it. The question then becomes: can we give a purely physical account of the nature of the self? Put differently, can we locate the self within the framework of physicalism in favor of the identity of the mental with the physical? Before we answer these sorts of questions in some detail, we need to see the full package of the strong physicalist's story which I primarily consider as a direct objection against Lowe's TS. In this respect, Frank Jackson's SM is a good representative. I now turn to a brief exposition of Jackson's SM.

6.2 Jackson's Serious Metaphysics (Location Problem)

In his *From Metaphysics to Ethics* (1998: Ch. 1),[119] Jackson tells a story about what he calls

---

[119] Frank Jackson (1998: Ch.1). In the philosophical literature, there are objections or counter-examples raised against different formulations of the theory of the self. Some of these objections include multiple personality disorder, split-brain patient, etc. But my interest here is not to revisit those familiar objections/counter-examples. Rather, I would like to see the issue raised here from a fresh angle (though by 'fresh angle' I do not pretend that my approach is entirely original) by formulating Jackson's SM as an objection against Lowe's TS and then give a rejoinder to Jackson's objections. For discussions surrounding the common objections against the theory of self see: Dennett (1991: Ch. 13); McGinn (2000: 157-174); for analysis and rejoinders see Lowe (1996: 26-7, 29-30, 40, 184, and 195); Swinburne (1986: 145-160). Recently Bayne, in his paper entitled "The Unity of Consciousness and Split-Brain Syndrome" argues that the disunity

SM within the framework of strong physicalism, according to which the physical nature of our world exhausts its entire nature. By SM, Jackson refers to a metaphysics that maintains that we can attain comprehension of the nature of things without drawing up big lists. That is, to understand the features of the world, all that we need is a limited number of basic ingredients. Even if for Jackson, metaphysics is about what there is and what it is like, this does not mean that metaphysics is concerned with any old shopping list of what there is and what it is like. Rather, as Jackson suggests, metaphysics seeks a naturalistic account (e.g., purely physical) of some subject matter—whether that turns out to be the mind, semantic properties, or everything—in terms of a limited number of more or less basic notions. So for Jackson, SM is the investigation of where these limits should be set. Thus, Jackson claims that by its very nature, SM continually faces the location problem. Because the ingredients are limited, some putative features of the world are not going to appear explicitly in some more basic account. The question is: whether putative features nevertheless figure implicitly in the more basic account. As Jackson claims, since SM is discriminatory, at the same time claiming to be complete, a host of putative features of our world either must be eliminated or located.

By location problem, Jackson refers to an attempt to locate some entities within the framework of physicalism. But Jackson understands that some entities such as semantic properties, *de se* content, mind, etc., do not seem to fit within the framework of physicalism. Thus, we face the location problem. However, Jackson claims that one way to solve the location problem is to embrace what he calls: the 'entry by entailment thesis,' according to which an entity is entailed by a basic physicalist story iff it is entailed by the basic constituent of that story. Moreover, such an entailment would enable us to identify what appears explicitly in an account from what appears implicitly in it. For Jackson, in this regard, science plays a pivotal role. For example, Jackson claims that science tells us about the basic constituents of tables, chairs, etc., without mentioning solidity. That is, science

_____

models of the split-brain are highly problematic and that there is much to recommend a model of the split-brain—the switch model—according to which split-brain patients retain a fully unified consciousness at all times." 'The Journal of Philosophy,' v. 6, (June 2008), 276-300. I will discuss Bayne's article in chapter eight.

tells us that a dry object like a table is an aggregate of molecules held by a lattice-like array by intermolecular forces. Such intermolecular forces keep the molecules from encroaching on each other's spaces. In such a scientific story, even if we have not come across the word 'solidity,' from this it does not follow that nothing is solid. Since objects like tables are solid, solidity gets location or place in the molecular story about our world by being entailed by that story.

Jackson gives three examples: (1) for Jackson density is not a property that is over and above mass and volume, instead density is entailed by mass and volume; (2) solidity is not an additional feature over and above the way a lattice-like array of molecules tend to repel each other; (3) in the same way, to say that Jones is taller than Smith is not something that we can take as a feature of how things are, which is additional to Jones's being five foot and Smith's being four foot nine, if in fact Jones is five foot and Smith is four foot nine.

So for Jackson, if the features of our world are entailed by the account told in terms favored by the metaphysics in question, it has a place in the account told in the favored terms. In other words, the one and only way of having a place in an account told in some set of preferred terms is by being entailed by that account. In the same vein, the psychological appears in the physicalists' account of our world iff that account entails the psychological account of our world. Hence, as Jackson states, the distinctive claim of physicalism is that "the world is entirely physical in nature, that it is nothing but, or nothing over and above, the physical world, and that a full inventory of the instantiated physical properties and relations would be a full inventory simpliciter," (1998: 9). Jackson also claims that an obvious way to approach completeness in physicalism is to assume that there is no independent variation. This suggests that we should look for a suitable supervenience thesis to capture the sense in which physicalism claims completeness. In light of the distinctive claim of physicalism, variation in the nature of a world independently of variation in the physical nature of that world is impossible.

Jackson also claims that egocentric or *de se* claims have owners. That is, such claims are about how things are with us. For example, I might believe that there are lions near me.

Moreover, egocentric claims have a perspective or point of view that is built into them. So they can be produced from the perspective of the holder of the belief. But for Jackson, this is as far as it goes. That is, since the strong physicalist story does not accommodate in any direct way notions like 'egocentric claims', 'one's point of view,' etc., Jackson claims that the best way to approach such notions is by giving a perspective-free account of our world. We can do this, according to Jackson, by reducing egocentric or *de se* content to the non-egocentric or *de dicto* content. As Jackson puts it, "…claims and beliefs about how things are with the claimer or believer can invariably be translated in some way or other into how things are with the world," (1998: 19). In short, we talk about how things are with the world as opposed to how things are from one's own perspective.

Finally, Jackson asks us to imagine a story in which ϕ be the story as told in purely physical terms. Jackson claims that such a story is true at the actual world and all the minimal physical duplicates of the actual world. But such a story is false elsewhere. The story imagined here is a hugely complex, purely physical account of our world. Jackson also asks us to imagine the ψ be any true sentence which is about the psychological nature of our world. By this, Jackson means that if supposing that there is such a true sentence, then it can only become false by things being different psychologically from the way they actually are. That is, every world at which such sentence (ψ) is false differs in some psychological way from our world. As it were, the idea is that ψ counts as being about the psychological nature of our world. The reason for this is that making ψ false requires supposing a change in the distribution of psychological properties and relations. In light of this, if we grant that any world which is a minimal physical duplicate of our world is a duplicate simpliciter of our world, then every world at which ϕ is true is a duplicate simpliciter of our world, and so a fortiori a psychological duplicate of our world. However, every world at which ϕ is true is a world at which ψ is true—that is how Jackson says that ϕ entails ψ (1998: 25).

I now turn to my formulation of Jackson's SM as an objection against Lowe's TS.

6.3 *Jackson's SM against Lowe's TS*

Remember that Lowe's TS, *inter alia*, is a claim about the distinction that holds between the self and the physical body that embodies it. But as we briefly saw in § 6.2 above, making such a sharp distinction between the self and the physical body, is unacceptable for philosophical naturalists who advocate strong physicalism. Thus, given the physicalist story as told by Jackson, it seems clear that the strong physicalist finds it difficult to embrace Lowe's TS. The strong physicalist equally finds it hard to endorse various sorts of mental phenomena, namely belief, desire, fear, hope or, in short, intentional states of the self, unless any such states are understood/described in a purely physical way. For example, Jackson argues this point explicitly elsewhere.[120]

What sort of a purely physical description can a strong physicalist give for Lowe's TS? This is precisely the answer that we want to know in light of Jackson's SM. In order for Jackson to succeed here, it seems decisive for him to be able to show us how we can get around notions like the self and still make sense of our experience of mental states (e.g., I feel a pain). But why is doing this important for a physicalist? It is important for at least four main reasons: (i) dualists think (or Lowe's TS entails) that intentional states as well as other mental states are states of a subject or states of their owners. If this is right, then it follows that mental states fail to exist on their own unless they have owners, namely selves or persons. (ii) Since mental states are not the same as physical states as shown in § 6.2, the strategy (so it seems to me) that the strong physicalist needs to put in place is the one that shows why the dualist view, both on mental states and on the nature of the self, is mistaken or even false. (iii) Since for Lowe, the self is an agent that is capable of doing intentional acts, the strong physicalist must also account for human agency from a purely physicalist perspective. (iv) The dualist claims that the self has first-person knowledge about oneself (e.g., the self knows that certain thoughts, experiences and actions are one's own). Since the strong physicalist denies first-person perspective, he/she should suggest a better way to understand claims made from first-person perspective via using first-person pronoun 'I' as

---

[120] For example, see Jackson's view on mental states in Michael Smith (ed.), "Consciousness," (2005: 310-333).

discussed in §§ 6.1 and 6.1.1. If the physicalist can come up with a better alternative account of (i)-(iv), then Lowe's TS can be shown to be either a thesis that should be rejected altogether or one that needs to be revised in a way that suits Jackson's SM.

To do this, as Jackson argues above, there seems to be three main options that are available: (1) locate the self or person within the physicalist ontology via a process Jackson calls, the 'entry by entailment' thesis according to which, an entity is entailed by a basic physicalist story iff it is entailed by the basic constituent of that story; (2) eliminate the self if it is not locatable within the physicalist ontology; (3) refuse from the beginning, on *a priori* or nonexperiential ground, to select the self as one of those ingredients to be included in the physicalist ontology. Notice here that Jackson's SM is discriminatory, that is SM can set a limit as to what entities are to be included and what others are to be excluded from the physicalist ontology. What then follows from the above options set up by the physicalist? Here are some of the things I think would follow. For example, if (1) obtains, then the dualist claim that the self is distinct from the physical body or any part of it that embodies it will be undermined. This is because locating the self within the physicalist ontology necessarily requires the obliteration of the distinction that Lowe thinks to hold between the self and the physical body that embodies it. Likewise, if (2) obtains then, the same result we saw in (1) will be true, that is, if the self is eliminated, then there will be no distinction that Lowe's TS implies between the self and the physical body that embodies it. In this case, we need to say goodbye to Lowe's TS. If (3) obtains then, the physicalist will have no reason whatsoever to bother himself/herself with thinking about the entity that has not been part of his/her ontology in the first place. So Lowe's TS simply remains a nonissue for the physicalist.

In Jackson's case then, of the three options listed above, the one that is important for the physicalist to explain away the self is to use option (1). In light of this, I do not take Jackson attempting to use options (2) and (3) above to explain away the self being distinct from the physical body or any part of it. This is because Jackson acknowledges that we have beliefs and also make claims with respect to how things are with us. For example, I might believe or assert that there are deers near me. Here the belief as to the presence of deers near me is

produced from my own perspective as a holder of such a belief. So we can at least say that Jackson's SM prima facie does not eliminate the self from the physicalist ontology. Instead Jackson is looking for a way to locate the self within physicalism. Thus, I suspect that the physicalist at this point might want to challenge Lowe's TS on the basis of (1). If the physicalist could do that, then Lowe's TS will be undermined. Given Jackson's SM, there are two main ways to do that:

First, the physicalist employs a third-person description strategy whereby first-person assertions or *de se* claims are deprived of having a first-person reference in favor of understanding such claims to be referring only to facts about the world. In other words, the physicalist can reduce *de se* claims, as Jackson tells us, to *de dicto* or non-egocentric claims. For instance, a claim like I myself wrote a letter is irreducibly about me, in a sense that I have the first-person knowledge of I myself writing a letter or my own experience of writing a letter. But the physicalist would say that such is not the case. Instead, the claim I myself wrote a letter is simply the physical property of writing a letter and nothing beyond. In other words, even if *de se* claims entail first-person perspective, in a sense that such claims are made from the perspective of the claimer, the physicalist story, as Jackson argues, is a "perspective-free account." So it seems then that the strong physicalist could easily explain away first-person perspective. Thus, in giving this sort of reductive analysis of *de se* claims, the physicalist can avoid the notion of the self or person. Recall here that Lowe's theory of the self argues that the notion of the self is captured by the first-person pronoun 'I' or an indexical "I". But if the physicalist's argument here is correct, then Lowe's theory of the self faces a serious challenge. This is because Lowe's TS implies that there is such a thing as a first-person perspective (for example, self-knowledge).

Second, the physicalist claims that the truth or falsity of egocentric or *de se* claims and beliefs supervene on the *de dicto* story about the world. That means *de se* claims and beliefs have no independent role whatsoever unless seen in the light of the *de dicto* claims. But such an assumption does not seem to be compatible with the first-person perspective scenario. We saw this earlier in § 6.1.1., in our discussion of self-knowledge. That is to say that what a self knows about oneself has to do with knowing that certain thoughts,

experiences and actions are one's own. But according to Jackson's notion of supervenience, variation in the nature of the world independently of variation in the physical nature of that world is impossible. What grounds this notion of supervenience is the conjunction of the following two theses: (*B*) "any world which is a minimal physicalist duplicate of our world is a duplicate simpliciter of our world" (1998: 12); (*B**) "any world which is a minimal physical duplicate of our world is a psychological duplicate of our world."(*Ibid*: 17).

The implication of such a physicalist supervenience thesis taken in light of (*B*) and (*B**) above for Lowe's TS seems to boil down to this: that nothing can be true of the self, which is not already true of the physical body or any part of it. That means that whatever is true of (*B*) necessarily must be true of (*B**) as well. If the physicalist is correct here, then it seems to follow, despite the dualist claim to the contrary, that the psychological properties are identical with the physical properties. But notice that Lowe's TS implies that the self is an emergent simple substance, that is, the self is ontologically distinct from the physical body or any part of it, such as the brain. Moreover, the self as an emergent simple substance is neither reducible to nor divorceable from the physical body that embodies it. But this does not seem to be the case, if Jackson's supervenience thesis above is the case.

Following Jackson then, the strong physicalist objection against Lowe's TS can be summed up as follows: if (*B*) above is true, then every world whereby the purely physical story told in the actual world is a duplicate simpliciter of our world. From this it also follows that (*B**) above is true: every world which is a minimal physical duplicate of our world is a psychological duplicate of our world. This means that for the strong physicalist, every world wherein the purely physical story told in the actual world is true, is a world wherein the story told about the psychological nature of our world is true. This again shows that the purely physical entails the psychological nature of our world. As Jackson puts it:

> What about sentences with egocentric content? We can think of a purely physical story as being in part about the physical nature of an individual as well as being about the physical nature of a world-thus the story might include…that I myself weigh 75 kilos-and by an obvious extension of the argument in the text, the physicalist must hold that the relevant egocentric psychological story about, say, me, is, if true, entailed by the relevant purely physical story about me as well as about the world (*Ibid*: 25).

Hence, for Jackson, it seems that the only way we can make sense of the notion of the self is by locating it within the framework of physicalism as stated under option (1) above. What then can we say in response to Jackson's SM? I now turn to that discussion.

6.4 *Critical Reflection on TS and SM Models*

The two different models, as presented by Lowe's TS and Jackson's SM, clearly lead us to two different conclusions as to the nature of the self. In a nutshell, Lowe's TS simply rejects any attempt to identify the self with the physical body that embodies it. Moreover, even if the self is a physical thing in virtue of possessing physical characteristics as discussed in §§ 6.1, 6.1.1, 6.1.2 and 6.1.3, for Lowe, mental properties which the self bears are not physical and thus, are not reducible to something physical. By contrast, Jackson's SM allows us to identify the self or person with the physical body simply because that is what best suits the scientific account of our world (e.g., Jackson's SM seems to verify this point in § 6.2). To put it differently, we do not need first philosophy to settle the issue of the nature of the self. Instead, we need to stick to the scientific account of our world. This is because it is only via science that we come to a good grasp of the nature of the self.[121] For example, given Jackson's SM, our world is such that only one sort of properties are exemplified (e.g., physical properties). Hence, the non-physical mental properties which Lowe's TS implies must undergo some sort of reductive analysis and be identified with the physical. In short, such is Jackson's way of arriving at a true scientifically justifiable picture of our world (Cf. Moreland 2008: Ch. 1).

So the key question is: which model best fits our ordinary conception of the self? By 'ordinary conception' here, I refer to what people in general think is obvious or central about their own experience as selves or persons. Of course, each side of the debate can favor its own conception of the self. So in that sense, the answer for the above question can go both ways (e.g., the dualist way on the one hand, and the physicalist way on the other). But obviously, we cannot affirm both answers for the simple reason that the answers we

---

[121] For example, see also Jackson's view on science in Michael Smith (ed.), "Consciousness," (2005: 310-333).

get from each side negate each other. So the question remains: how should we address such a problem? The best thing to do here is to examine the reasons given by each model on the nature of the self. If we accomplish such a task, then we can be in a better position to adjudicate over which model gives a plausible account of the nature of the self. In doing so, we can also have a better grasp of which model best reflects our ordinary conception of the self.

But the strong physicalist might protest here by saying that we do not need to be concerned with our ordinary conception of the self or person. To know about ourselves, all we need to do is to pay attention to what scientific theories currently available say about our world and the place we have in it. The main reason then for such disapproval of our ordinary conception of the self is its incompatibility with currently accepted scientific theory, according to which the notion of the self is analyzed in terms of third-person description. I will say more on this very objection at a later time. I will return to this issue again in chapter seven. But for now let me give one reason as to why I think our ordinary conception of ourselves, as persons, is pertinent in our overall effort to understand what it means to be a person.

Suppose that I get seriously sick, and thus I go to the hospital to see a doctor. Insofar as I am able to communicate, the first thing the doctor does is to ask me to explain to him/her how I feel. Of course, I may misconstrue and thus, misreport how I feel. But the report the doctor gets from me is very important to his/her understanding of the health problem I have, which in turn helps the doctor to decide what kind of treatment I should get. So the report I give to the doctor on how I feel is not a theory I have over what kind or how many sorts of feelings I have. Instead my explanation of how I feel is a direct report that I give to my doctor from my own first-person perspective. Here, my experience of how I feel has, in the words of Sydney Shoemaker, "immunity from error through misidentification."[122] For example, if I experience pain in my abdomen, then it is hardly plausible to think that I

_____

[122] Sydney Shoemaker, "Self-Reference and Self-Awareness", *Journal of Philosophy* 65 (1998): 555-67., as quoted in Lowe, (2000: 267). Also see Shoemaker (1996: 15, 196-7, 211).

misidentify such a pain experience with someone else's pain experience. Of course, it may be the case that I may report wrong things about how I feel, but that *per se* does not invalidate the fact that there is such a thing as the first-person perspective, such that I necessarily know about my own pain experience. The point here then is not the accuracy of the report I give to a doctor on how I feel, but whether I could give a report of how I feel (whatever truth value it has) from my own first-person perspective.

Here let us also consider the 'phantom' pain phenomenon. Sometimes people with an amputated leg claim that they still experience pain in the amputated leg. Imagine saying to such people that they simply must be kidding and nothing they say is true. We say this because it seems ludicrous to believe that there can be such a pain experience in the absence of a leg that has already been amputated. Such discussion can be controversial but getting into that is not my wish here. But the question is: on what basis can we then deny such reports of the 'phantom' pain, since the 'phantom' pain experience itself is being reported from someone's first-person perspective? The answer clearly is that there is not any plausible basis for a third party to deny someone's claim of the 'phantom' pain. The lesson here, is that as Lowe argues, the capacity for self-reference and also knowing certain experiences to be one's own are some of the fundamental features of our experience via which we develop our ordinary conception of what it means to be a self or person.

So the bottom line here is that contrary to the strong physicalists, our ordinary conception of our experience of things from our own first-person perspective, and as a result of that, the report we give on our experience is not irrelevant at all. In fact, the claims we make from first-person perspective do not depend on scientific theories of one sort or another, for the simple reason that such claims resist third person analysis. I will say more on this at a later time. Thus, to deny first-person perspective because of its incompatibility with a scientific theory of some sort and because of that to claim that we should redefine the notion of the self in terms of third-person description is a *non-sequitur*.

The strong physicalist can still try to suggest other counter-examples, like in the case of infants or babies who cannot explain how they feel, and people who go into a deep coma or

become temporarily unconscious, and so must depend on their doctor's third-person analysis of their situation. Thus, such scenarios can be defeaters for the preceding discussion on first-person perspective. In response, we can say that the assumption here is misguided. In the case of both infants and people who are in a deep coma or are temporarily unconscious, their own first-hand experience is left unarticulated due to different reasons. In the case of infants or babies, we can say that they have not yet developed their capacities of speech to engage in verbal or linguistic activity. But from this it does not follow that infants do not experience things from first-person perspective. This is because the absence of an immediate ability to verbally communicate, or even not being aware of things from the first-person perspective, does not rule out the capacity babies have to communicate, for example, by crying, becoming restless, not wanting to eat, etc., to indicate that they are not feeling well. Of course, things such as crying, lack of appetite, etc., are behaviors we observe from the third-person perspective but this by itself still does not show that babies themselves do not have their own first-hand story to report even if we know that they cannot verbally report at the moment. If I am right here, then for infants or babies, it is just a matter of time until they reach a certain age level where they can verbally express themselves and their experiences from a first-person perspective. I will say more on issues related to this point in chapter eight.

We can offer a similar argument when it comes to those who experience a deep coma or become temporarily unconscious. Due to a serious medical condition or because of an accident, people might not be able to articulate their own experience. But it is fallacious to conclude from such a scenario that they lack a first-person perspective regarding their experience. This is because persons have first person perspective, independent of some situations that might prohibit them from exercising it. For all we know then, people at such moments cannot verbally utter their experience nor do they seem to be aware of what they experience. Of course, no one can say anything whatsoever on behalf of what their experience would be like other than the people themselves. We can only wait and see what these people would say regarding their experience after they gain back their consciousness. Therefore, we can safely say that the physicalists' attempt to replace first-person perspective by third-person perspective is deeply implausible. If I am right, then first-person

perspective necessarily belongs to oneself. Thus, even if the doctors analyze the problems of both the infants and the people who are in a deep coma or temporarily unconscious from the third-person perspective, what the doctors still fail to know, is as Nagel argues, 'what it would be like' to be sick or be in a deep coma or to temporarily become unconscious, etc., from the perspective of people who go through such experience (see for details Nagel 1986).

In light of what has been said so far, it seems to me that Jackson's claim that we can do away with *de se* or egocentric claims by reducing them to *de dicto* claims or facts about the world is problematic. This is because the process of reducing *de se* claims to *de dicto* claims necessarily brings the loss of first-person perspective. But since, as we saw above, the first-person perspective is irreducible, it follows that we cannot capture the meaning of *de se* claims via *de dicto* claims. Contrary to Jackson then, the notion of *de dicto* claims cannot play the role of the *de se* claims without thereby leaving out the first-person point of view. As Lewis argues, "…some thought is egocentric, irreducibly de se, and then its content cannot be given by the propositions whose truth is relative to nothing but worlds; for propositions do not discriminate between inhabitants of the same world," (1986: 55). As Lewis further remarks, there is no difficulty in understanding how different sorts of egocentric experiences are irreducible; unless we stubbornly insist that such experiences cannot be what they seem (*Ibid*. 125).[123] In light of what I have said so far, Lewis is correct in telling us that we can take various *de se* thought or belief instances at face value, that is without reducing *de se* claims to *de dicto* claims. If this is right, then Jackson is mistaken in his claim that the truth or falsity of *de se* claims are fully determined in the light of the *de dicto* claims.

But Jackson could respond to such a rebuttal by saying that the physicalist story (which is based on science) is a 'perspective-free' account and thus, the physicalist should not be concerned to retain the first-person perspective in a way that the dualist wants. But Jackson's response faces its own problems. First, it seems to beg a question in that if the

---

[123] See also Lewis's "Attitudes De Dicto and De Se", *Philosophical Review*, 88 (1979): 513-543

physicalist account of the world is a 'perspective-free account' of our world, then isn't the physicalist assuming a 'perspective' to tell us about the physicalist story being a 'perspective-free account?' Second, since science *per se* neither establishes nor can tell us about the nature of reality (as I will try to show at a later time) without some sort of philosophical assumptions first put in place, it remains both entirely unclear and also controversial to invoke a 'perspective-free account' of the world as Jackson states it.

But there is another serious problem with Jackson's claim about a 'perspective-free account'. Suppose that the physicalists think that their position, physicalism, is the accurate description of the world we live in and who we are as human beings. Suppose further that the physicalists formulate a theory known as strong physicalism (e.g., as Jackson's own theory SM shows). At this point, the physicalists obviously want to convince their opponents of the plausibility of their physicalism. I would assume this is also true of not just physicalists, but in general anyone who defends some sort of theory in philosophy or science or any other discipline. If there is a position one argues for, then how can we divorce the perspective from which it is argued? If indeed as Jackson claims that physicalism is a 'perspective-free account', then what does physicalism amount to as a theory? How can physicalism be a position that the physicalists argue for if the position itself is empty of perspective? Jackson might obviously suggest as a response here the need to reduce egocentric claims to non-egocentric claims. But as we have already seen, that kind of reduction will not help the physicalist. The only alternative the physicalist seems to have here, I think, is either to drop his claim on the so called 'perspective-free account of the world' or bluntly face the problems raised above. But none of the alternatives seems impressive, to say the least. It seems then fair to say that Lowe's TS model in emphasizing the first-person perspective as one of the fundamental features of the nature of the self or person poses a formidable challenge to the strong physicalists' denial or attempt to reduce first-person perspective to third-person description or perspective.

The other way Jackson's SM challenges Lowe's TS is by explaining away the self via supervening the truth or falsity of egocentric or *de se* claims and beliefs on the *de dicto* story about the world. That means that *de se* claims and beliefs do not have an independent

role to play unless they are seen in the light of the *de dicto* claims. But such an assumption is incompatible with the first-person perspective scenario we discussed above. Since I have already said enough as to why reducing or leaving out the first-person perspective from one's ontology is problematic, at this point, I only want to respond to one crucial aspect of Jackson's supervenience claim. Remember, for Jackson, supervenience is the conjunction of two claims as identified earlier as (*B*) which states that any world which is a minimal physical duplicate of our world is a duplicate simpliciter of our world and (*B**) which states that any world which is a minimal physical duplicate of our world is a psychological duplicate of our world.

Once the supervenience relation between *de se* claims (made from first-person perspective) and *de dicto* story about the world are understood in light of (*B*) and (*B**) above, it seems that the dualist has to think about what the causal history would be like between the physical and the psychological. Remember that dualists attribute causal power to the self without which the self would lose one of the decisive features of selfhood. Dualists also argue that the self is a bearer of mental properties which are non-physical states. Since *de se* claims imply mental states (e.g., I am afraid of bandits near me) which are causally autonomous, Jackson's supervenience claim here, if true, creates a problem of causal exclusion. For example, Kim argues that the problem of causal exclusion arises from the "supervenience" argument and it is the main problem of mental causation (1998: Chs. 1-2). One of the things that make the exclusion problem distinctive is that it strikes at the very heart of physicalism. Hence, the basic problem of mental causation is concerned with giving an account of how it is possible for the mind to exercise its causal powers, given a world that is fundamentally physical. So the problem of causal exclusion boils down to showing whether a mental cause is possible, since every physical event that has a cause has a physical cause. The question then becomes, as Kim remarks: what place does mental cause have in a world in which only physical cause is responsible to bring about physical event? Or if we acknowledge that P has a physical cause P^* at t, then the question to ask would be: given that P has a physical cause P^*, what causal work is left for M to contribute?

The biggest worry for the physicalists then, is that in a world where only physical causes are acceptable, leaving room for the non-physical causes or mental causes would have a troubling result for physicalism. As Kim remarks, if mental event m, occurring at time t, causes physical event p, and if this causal relation holds in virtue of the fact that m is an event of mental kind M and p an event of physical kind P, does P also have a physical cause at t, an event of some physical kind N? Here, to deny that p has a physical cause at t, yet at the same time to grant that only mental event m taking place at t as a cause of physical event p would be a violation of the causal closure of the physical domain, according to which no physical event has a non-physical cause. As I see it, at this point, those philosophical naturalists who accept the causal closure principle as stated above will unite in rejecting top/down causation. That means that given the physicalist story, non-physical mental causes do not operate in the physical world. Thus, as Kim argues, the physicalists can avoid overabundance of causes, for example, if they take P to be a cause of P^* and M supervenes on P, then M^* supervenes on P^*. Hence, given the physicalists' account, the M-to-M^* and M-to-P^* causal relations are only apparent, arising out of a genuine causal process from P to P^*. I think Kim's conclusion here directly resonates well with Jackson's supervenience claim—grounded in (B) and (B^*) above.

The question remains: what implications does Jackson's move to supervene *de se* claims upon *de dicto* claims have for the nature of the self? As we saw earlier, if the *de se* claims are rooted in the first-person perspective in a sense that such claims are produced from the point of view of the holder of a certain belief, then such claims clearly have a mental content. Moreover, for the dualist, claims with a mental content are causally relevant. But as the strong physicalists argue, if the mental cannot cause the physical, then every causal phenomenon has to be reduced to physical cause. In this sense, strong physicalism is inherently reductionistic. For example, if mental properties which the self or person bears, lack genuine causal power and appear to be having only an apparent causal power, then by strong physicalist's reasoning, the psychological and the physical properties must be identical. That means that we can only talk about physical causation as opposed to mental causation. Why should we think that is the case? Since our mental states–our beliefs, desires and intentions–have causal effects in the physical world, to deny that mental states

215

have a genuine causal power is counterintuitive at the least. Moreover, the physicalists' causal closure claim, as stated above, is problematic for the simple reason that it begs a question. This is because the strong physicalists assume that without reducing mental properties to a physical domain, mental properties are irrelevant to establish causal relation between the mental and physical states.[124]

So even if the supervenience relation between the *de se* claims and the *de dicto* claims holds, it will not be a solution for the physicalists to establish their claim that everything must be described on the basis of physics. This is because giving the physical description of the supervenience relation as stated above, per se, does not answer the key question, why in the first place such a supervenience relation exists between the *de se* claims and the *de dicto* claims. Thus, the supervenience relation here simply remains a problem yet to be solved. Insofar as I can tell, Jackson's SM has not solved it. But Jackson might respond to the

---

[124] Lowe (2008: 25-40; Chs. 2-3), looks at several arguments surrounding the causal closure of the physical domain. In my opinion, Lowe in his book raises devastating challenges against physicalism in general, and strong physicalism in particular. For lack of space, I will not go into those arguments. However, I will say something briefly about the main ideas of the arguments. Lowe evaluates different versions of the causal closure arguments and shows why he thinks that it is possible for the psychophysical dualist to be open to the stronger causal closure principle which will not invite the transitivity problem. According to Lowe, a persuasive case can be made with respect to taking mental states to have a full causal power, in a sense that mental states can cause physical events. But such an assumption need not lead one to think that mental states are not themselves a product of prior physical evolution. In fact, part of Lowe's main point is to argue for the thesis that mental states are themselves a product of prior physical evolution, yet, Lowe makes it clear that the distinction between the mental states and the physical states needs to be maintained. Of course, physicalists would not want to embrace the claim that non-physical mental states cause physical events. For the physicalists, such causal relation between the mental and the physical would amount to violating the physical closure principle. But, on the other hand, if the physicalists' physical closure principle is true, then the dualists claim that mental states can cause physical events would be false. But Lowe argues that there is no good reason to let the physicalists get away with their causal closure claim for the simple reason that it is perfectly possible to argue that ultimately, mental causation can be shown to fall within the framework of physical evolution. If this is true, then as Lowe claims, we can understand mental causation to be linked with the physical causal history. As Lowe puts it, "it is possible…that every physical event has a set of wholly physical causes which are collectively causally sufficient for the occurrence of that event-and yet for it also to be true that some physical event, P, has a non-physical event or state, M, amongst its causes (without envisaging this as involving the causal overdetermination of P). This is because M itself may have a set of wholly physical causes which are collectively causally sufficient for its occurrence. If M is a cause of P, then, by transitivity of causation, all of those physical causes of M are also causes of P- and clearly they may form a subset of wholly physical causes which are collectively causally sufficient for the occurrence of P." (p.26). Such an assumption as laid out above, is perfectly consistent as Lowe argues with (i) the self or person being distinct from the physical body or any part of it, such as the brain; (ii) the self or person being an agent capable of engaging in intentional acts, some of which bring about physical outcome. If Lowe is right, then the strong physicalist's claim that denies a non-physical mental cause for physical event must be rejected.

above remarks by saying that his physicalism is only a claim about the actual (our) world to the effect that its physical nature exhausts everything about it. In light of this, Jackson may even grant that the above mental causation scenarios might be acceptable, given that such mental causations are not taking place in our world. So insofar as our world is concerned, autonomous non-physical properties cannot be instantiated.

In response to Jackson, I have two things to say. First, Jackson's claim that any physical event is not outside of the physical domain and everything can be explained on the basis of physics is problematic. This is because such is a strong philosophical claim which can hardly be accommodated within the physicalist domain. That is, the statement ''everything can be explained on the basis of physics'' itself cannot be explained on the basis of physics or science. That means that the statement itself cannot live up to its own scrutiny. If this is right then, to make a claim that a complete physical explanation of everything is possible is self-defeating. Second, contrary to Jackson, for all we know, our world is such that non-physical mental states can cause physical events. For example, if I believe that it will rain this afternoon, then I might take an umbrella with me to school.

But if mental states do not have a genuine causal power, then they are merely epiphenomenal or causally impotent. This means that there is no top/down causation that is from mental to the physical. Instead, for the physicalists there is only bottom/up causation (e.g., from the physical to the mental). But denying top/down causation, as the physicalists do is deeply flawed, given what we know about intentional states and our own first person awareness of them. Thus, no plausible physicalist case is made that refutes Lowe's TS. However, an objection can be raised against these remarks as follows. First, one one may say that our personal experience of intentional states as causally efficacious is just an illusion. Second, what I considered 'implausible' above is what a physicalist will consider plausible. So, my case against physicalism remains unpersuasive.

My response to the first objection rests on the semantics of the word 'illusion' which as I see it, can be understood both epistemically as well as metaphysically. Taken in the former sense, X is said to be an illusion just in case for all we know, X is not what it appears to be.

Taken in the latter sense, X is an illusion just in case there is no fact of the matter as to the very existence of X. Of the two senses, the latter one is the strongest. In light of this, if the causal efficacy of intentional states such as our desires and beliefs are said to be an illusion, then the most familiar part of our life is not what we normally think that it is. To see this, consider this scenario. When I feel thirsty, I open my refrigerator and look for juice/Coca-Cola or water. In this case, I believe that there is water in the refrigerator. In drinking water, my desire is to quench my thirst. Before I cross the road, I make sure that the road is clear from the traffic; otherwise, I get hit or killed. Examples of these sort can be multiplied in their thousands. The point is that in all such cases, we take for granted that we act on the basis of our desires or beliefs to explain our actions as we experience the world around us, as well as interact with other fellow humans.

Of course, this does not mean that our experience is incorrigible or indefeasible. But from this it hardly follows that our intentional states have no causal input into our actions. In this case, our realism in the causal efficacy of intentional states is not contingent on us first refuting the 'illusion' objection. So it remains unclear what motivates a defender of the causal efficacy of intentional states to take up the task of refuting the illusion objection before she establishes her case. More importantly, if the 'illusion' objection is taken in its metaphysical sense, then the burden of proof squarely rests on the one who says that intentional states do not even exist, to give us concrete evidence that substantiates such claim. In this regard, if there is one thing that the last five decades debate in philosophy of mind has shown, it is how difficult it is to establish the 'illusion' objection (see e.g., Kim, 2006; 2005). Hence, the 'illusion' objection against the causal efficacy of intentional states does not seem to pose any serious threat to those who defend it. My response to the second objection revolves around the notion of 'plausible' vs. 'implausible'. These notions are closely related to the notion of persuasion'. If one finds certain philosophical argument plausible (i.e., reasonable) then one most likely will find it persuasive. On the other hand, if one does not find certain argument plausible, then one most likely will remain unpersuaded by it. In light of this, the dualists find the physicalists' denial of the causal efficacy of intentional states implausible. On the other hand, the physicalists find the dualists' realism about the causal efficacy of intentional states implausible. So what is going on here? Is the

disagreement between dualists and physicalists over the causal efficacy of intentional states substantive or merely verbal? This is a metametaphysical issue which I have no interest to pursue for present purposes (see e.g., Chalmers, Manley and Wasserman 2009). That said, however, the plausible vs. implausible objection raised here cuts both ways in that it does not favor the dualists' position over that of the physicalists' or vice versa. So the only thing one can do in this case is to state one's position as best as one can. Although persuading someone is what a philosopher always hopes to do as she presents her case, it is not within her power to ensure that persuasion actually happens. We need to keep in mind that persuasion is person-relative. In philosophy, even after we present our very best argument in defense of our position, our opponent, if she wishes, could always find some reason to deem it as implausible. This is one reason why there is almost no issue in philosophy which is free from controversy. So I conclude that as it is in the case of the first objection, the second objection also does not seem to threaten the dualists' realism about the causal efficacy of intentional states.

6.4.1 *The Location Problem*

To see whether or not Jackson's SM model solves the location problem (as presented in § 6.2), it is important to answer a simple question such as this one: are there things for which we cannot find a place in the physical world? Put differently, is a purely physical description of everything on the basis of modern science possible? Given Jackson's SM, the physicalists would say yes. But the dualists deny such a positive response, yet they obviously grant that certain things can find their place in the physical world or can be described on the basis of physics. So, if the dualists want to rebut the physicalist's bold claim above, then he/she should point out entities that resist having their place in the physical world or resist the complete physical description of their nature. If the dualist could show this, then the physicalist's claim that every entity indiscriminately can find a place in the physical world will be undermined. For present purposes, I will raise only two points.

A. *Emergent Properties*

Jackson's SM model ignores fundamental aspects of the nature of the self. To see this point we need to recall that for Lowe, even though the self is a physical thing in a sense that the

self possesses physical characteristics, the self is a bearer of non-physical mental properties. So, one of the things that characterize non-physical mental properties is that they are emergent properties.[125] An emergent property, as Moreland argues, is a simple non-structural property unique to the "emergent" level of reality. Hence, an emergent property cannot be captureable, for example, in terms of the spatio-temporal, physical, causal relations among the states as well as parts at the lower level reality. Moreover, as Moreland further notes, an emergent property, being a new kind of property, is different from and not composed of the parts, properties, relations, and events at the subvenient level. If this is the right characterization of an emergent property, then this would put Jackson's SM in an awkward position. That is to say, an emergent property does not fit within the physicalist ontology. Thus, an emergent property cannot be entailed by physicalism as the basic constituent of that story. For example, phenomenal consciousness has a qualitative feel (for example, seeing yellow) in a sense that it is about ''what it is like to be sentient''.

Suppose then that I experience a sharp pain in the middle of my back. Suppose further that my wife looks at my grimaced face and asks me what is wrong with me. Will I then tell my wife to go to the hospital to ask the doctors to find out about my pain experience? The answer is clearly no. This is because, since I am the one who experiences such pain, I have direct access to my own pain experience which qualifies me to tell my wife how *hurtful* the pain in my back is. The pain in this case has a distinctive *quale* or felt sensation that constitutes my pain experience. If I am right, then all that the doctors can do is analyze what is wrong with my physical body from the third-person point of view. But giving an account of the experience of pain itself and its distinctive *hurtfulness* has nothing to do with physics and chemistry or the scientific analysis of the pain.

Of course, the strong physicalists would reject the above argument for at least two reasons. First, the physicalist might say that since mental states are identical with physical states, there is nothing to introspect or privately access as if mental states are distinctly recognizable. Second, even if mental states exist they are epiphenomenal, i.e., with no

---

[125] In this regard, see an excellent discussion in Moreland (2008: 1-7).

causal role whatsoever. Hence, there is nothing to introspect about causally impotent states. But such rejection is rooted in the physicalist's ''grand'' story about our world, which is entirely physical (see also Moreland 2008: Ch. 1). But the dualist can still argue that there is no good reason to endorse the physicalist rejection of mental states. In response, the dualist can run a similar argument as advanced above. That is, we can introspect our mental states and properties. Since mental states are properties that we exemplify, then mental properties stand to us as inner. But ''inner'' here should not be taken to mean spatially inside, since ''inner'' stands for property exemplification. Moreover, our mental properties are not accessible by a third party unless we ourselves want to make that to be known to others. In other words, we have direct unmediated access to our own sensations and conscious states (see also Chapter one § 1.1.1.1).

But none of this is true of physical objects or brain states. For example, my sensation of my car is not mediated by other sensations. This is a self-presenting property when instantiated by a subject, it presents both the intentional object (car) and itself to the subject. As Lycan, (who is not a friend of dualism) remarks, "to anyone uncontaminated by materialist philosophizing, the mental does not seem physical in any way at all, much less neurophysiological," (2008: 2). Lycan's remark here clearly strengthens what the dualists have been saying all along about the non-physicality of mental states. In light of all this, the physicalists' denial of mental properties being accessible privately to their owners flies in the face of our experience. Thus, physicalism fails to give us a complete physical description of the nature of the self that bears non-physical mental properties. *Ergo* emergent properties cannot be located in the physical world.

By contrast, we can give a complete physical account of things if, in this case, what we have in mind is structural property. Here by structural property, I refer to a property that is identical to a certain way of arrangement among the subvenient level. This means that unlike emergent properties, structural properties are not new kinds of properties but rather they are new arrangements of subvenient entities. Since there is no new kind of entity here, nothing stands over and above those entities that are already at the subvenient level. For example, Jackson's analogy on solidity, as stated in § 6.2, is a perfect example of structural

property. As Jackson argues, science tells us that a dry object like a table is an aggregate of molecules held together by a lattice-like array by intermolecular forces.

In such a scientific story, even if we have not come across the word 'solidity,' from this nothing follows to the effect that nothing is solid. Since objects like tables are solid, solidity gets location or place in the molecular story about our world by being entailed by that story. However, Jackson's analogy faces a problem here. That is, solidity is disanalogous to a *quale* as stated above. While there is a 'what it is like to feel' a pain (emergent property), the same cannot be said of solidity. From all this then, it follows that physicalism is ill-prepared to deal with emergent properties.[126] I will say more on the notion of emergence in chapter eight.

B. Agency

Jackson's SM model explains away the fundamental aspect of human agency. To get clear on this point, we need to remember that for Lowe, the self is by its very nature an agent that is capable of engaging in intentional actions, some of which bring about physical results. One of the difficulties we face then in discussing human agency, has to do with the question of whether human free will is compatible with determinism or not. Of course, one can deny that human agents possess libertarian or self-determining freedom in favor of holding that agents are only free to do as they want and are morally responsible for the choices they make. I explain the details of this argument in what follows. In light of this, one

---

[126] If the above discussion on emergent properties vs. structural properties is correct, then it would have great significance for those who advance arguments for consciousness as an emergent phenomenon that does not depend on the subvenient level in a way that structural properties do. For example, see Moreland's Consciousness and the Existence of God, 2008; Richard Swinburne, The Evolution of the Soul, pp. xii and chapter 10; Eric LaRock, argues for the irreducibility of consciousness on empirical ground, see his essay 'Is Consciousness Really a Brain Process?' *International Philosophical Quarterly,* v. 48, No. 2, Issue 190 (2008), 201-229. But philosophical naturalists of different stripes have different reactions on the nature of consciousness. For example, Searle takes consciousness as an emergent phenomenon, but locates it within the neurobiological processes (see e.g., Searle 1992: Chs. 4-5); McGinn argues that consciousness is a deep mystery (see e.g., McGinn 1999: Chs. 1-2); Churchland dismisses consciousness (see e.g., Churchland 1988), for Jackson's view see Jackson in Michael Smith (ed.) (2005). My own view is in line with Moreland and Swinburne that consciousness is an emergent property. Since the self is the bearer of consciousness, in light of the reasons we saw so far, it also seems hardly possible to find a place for consciousness in the physical world in a sense that the physicalists point out. I will say more regarding the emergence of consciousness in chapter eight.

can conclude that human free will is compatible with determinism.

For present purpose, we are only interested in understanding the notion of human agency, given that our world is all in all physical. By 'human agency' I refer to the self's capability to perform intentional actions. Notice that for human agency to obtain, the antecedent condition, in this case mental causation has to be possible. However, given our preceding discussion, we cannot find a place for mental causation in a world that is fundamentally physical. Then the question becomes: which model, SM or TS, fits our ordinary conception of human agency? Of course, the strong physicalist might say that such a question is an irrelevant one in relation to his/her main argument. For example, Jackson argues that redefining our ordinary conception of human agency is one of the tasks that his SM sets out to accomplish. As Jackson puts it, "there is nothing sacrosanct about folk theory. It has served us well but not so well that it would be irrational to make changes to it in the light of reflections on exactly what it involves, and in the light of one or another empirical discovery about us and our world," (1998: 44). So from Jackson's remarks, it seems clear that as a physicalist, he wants to reformulate the way human agency is supposed to be understood. For example, for Jackson our intuitions are the primary sources for our view of free action, determinism, etc.

But in order to have a correct understanding of what we claim to know on the basis of intuitions, there is further work to be done, which would include elucidating the possible situations covered by the words we use to ask questions regarding free actions, among others. At this point, the dualist legitimately may ask whether the pre-theoretical conception of human agency needs to undergo a sort of revision that the physicalist demands. For present purpose, I do not want to pursue matters related to Jackson's conceptual analysis. However, in light of his SM, the big question we need to answer is this: can we locate agency in a world that is fundamentally physical? To answer this question, let us briefly analyze the notion of human agency in light of strong physicalism on the one hand, and dualism on the other.

Since the SM model primarily is a claim that everything must be described on the basis of physics, the best we can get from such a model is a sort of human agency which is rooted in determinism. By determinism[127], here I have in mind a notion that refers to a family of views. Taken this way, determinism stands for the idea that everything in the world, human action included, is controlled by universal laws of nature. 'Psychological determinism' is another member of that family, according to which our desires, purposes, needs, etc., are motivators of our behavior. In light of this, if SM is correct, then for every event E that happens, there is an antecedent cause C such that given C, nothing else could have happened. That means that if determinism is true, then every human action, say for example, raising my hand in one of my classes to ask a question, is causally necessitated by prior events that already took place prior to my hand going up, which also includes events that existed before I was born. According to this model then, human agency remains fixed by the way the world is.

On the other hand, given the TS model, we can say that free human actions are not compatible with determinism. If human agents have real freedom, that is freedom undetermined by an antecedent cause of some sort, then it follows that agents have control over their own actions. If understood in this sense, libertarianism is a view that holds that agents determine their own actions freely. Unlike determinism as stated earlier, libertarianism rules out antecedent causes or factors as being responsible for an agent's choice of action. Of course, one's desires and beliefs are capable of influencing one's choice. But from this it does not follow that free acts are caused by antecedent factors in the agent. Hence, agents have control over their own will or at least have control to will to act. For example, if I am given a choice to live in the US for the next ten years or go back to my home country Ethiopia, nothing save for my own will determines that either choice is made. Since I am an agent, I can exercise my own causal power to do one alternative or refrain

---

[127] Hard determinists deny the existence of free will whereas soft determinists accept the free will but claim that it is compatible with determinism. I will not be concerned with either view here. There is also a version of determinism known as theological determinism, according to which if God knows in advance what free agents would do then such divine knowledge of the future actions of free agents makes human actions to happen out of necessity in which case, human agents are not free to do things out of their own free choice. This version of determinism will not be discussed in this chapter.

from doing it altogether.[128] According to the TS model, selves or persons are capable of bringing about effects that have not been predetermined by something other than the agent itself. So in the TS model, human agency is not fixed in terms of the way the physical world is.

Given the above discussion, the SM model faces serious difficulty in finding a place for agency in a world that is fundamentally physical. Here is why. First, since mental causation is not possible in a world that is fundamentally physical, it makes perfect sense to say that genuine agency is not possible either. Second, since the self is an agent and genuine agency requires causal autonomy, it follows that a complete physical world turns out to be unable to accommodate agency. This is however, only if the physicality of the world is taken in a strict physicalist sense. On the other hand, notice that Lowe argues that his TS is consistent with the following claim: *"Every physical event has a set of wholly physical causes which are collectively causally sufficient for the occurrence of that event—and rarely if ever is a physical event causally overdetermined,"* (2008: 20). Third, since mental causation is a species of an emergent property, and since an emergent property is not locatable in the physical world (for reasons we already saw), it follows that the SM model simply leaves out agency. By leaving out agency, the SM model faces another problem which I will discuss at a later time. So this leaves us with the TS model as a plausible alternative.

Jackson might reject the previous analysis of agency by saying that if we expand the extension of the term or concept we use to talk about 'free action' or use the phrase 'free action' to cover more cases than what is normally recognized, then it can be shown that free action is compatible with determinism. If Jackson is right, the objection is that folk conception of free action as stated in a libertarian sense will no longer be needed. As a result, the physicalist can say that free action amounts to actions which have been prefixed or predetermined to happen in accordance with physical law. In response, we may agree with Jackson in a sense that we can paraphrase concepts and words we use to mean slightly different things than the way we frequently use them. We may even expand the extension

---

[128] See for example, van Inwagen (1983); also see Lowe (2008: Chs. 6-10).

of what those concepts and words mean. But this is as far as Jackson's conceptual analysis brings us. This is because his conceptual analysis method turns out to be very problematic when it comes to moral responsibility. When people take responsibility for their actions, it seems to me that this is because there is a deeply entrenched intuition that informs them that they acted one way while still they could have acted the other way, had they chosen to do so. It is usually in light of this realization that people show true remorse for their bad actions; praise for their good deeds. Surprisingly enough, even children at an early age are capable of reasoning in this fashion. For example, children know that when they do something they should not have done, they will be held accountable for their actions. Such understanding may not be as precise as we like it to be, but the point is that when people reason out properly, they realize that for some of their actions they will be held accountable.

Suppose that John embezzles more than three billion dollars from a company he works for and reasons to the company's administrative office in these words: 'Folks, what I have done is something that has a long causal-chain that goes back to the time even before I was born. So I have no good reason to feel sorry, let me go home.' Who would buy John's reasoning? The answer is obvious, no one, unless there were a really good reason to the contrary. For this and countless other scenarios which have to do with human actions, the only plausible response would be to assume the libertarian view of freedom. So given a libertarian view of freedom, John was in complete command of himself when he did what he did. Only if we have such convictions about human free will, will we be justified in taking responsibility for our actions. But how could this be true if the world is such that only physical laws are responsible for causing anything that happens? How can we make sense of the very idea of taking responsibility for our actions? For example, it is very clear that we ungrudgingly take responsibility only if we believe that we have a causal power, such that the action performed is the result of an intentionally informed decision. As far as I can tell, Jackson's SM does not have any plausible answer for these sorts of questions.

If indeed every human action is the outcome of factors existing prior to an agent's acting in

a certain way to bring about some sort of result, then the very notion of agency is confused. On the contrary, if an agent is capable of bringing about effects by *willing* (as we discussed in § 6.2), then the physicalists' claim that a complete account of everything including human free will can be given on the basis of a fundamentally physical world must be false. It makes perfect sense then to say that human agency is hardly locatable in a world that is fundamentally physical. If I am right here, then it follows that Jackson's SM fails to locate some important things, one of which is agency.

6.4.2 Methodological Issues and Jackson's SM

As we saw previously, Jackson's SM cannot locate important things that define the nature of the self, namely emergent properties and agency. Of course, the list here is not exhaustive. But we have said enough to make our point that Jackson's SM leaves out important things mentioned above from the physicalist ontology. As we will shortly see, in leaving out things like emergent properties and agency, physicalism will face its own serious problem. It seems then that some of the main reasons for the physicalism's inability to locate emergent properties as well as agency within its ontology have to do with at least two things: (i) the rejection of first philosophy; (ii) the setting forth of science as the only credible and adequate methodology to account for the nature of things. When the strong physicalists (as Jackson represents them) stick to (i) they often fail to give a complete account of the nature of things on the basis of only (ii).

As Lycan remarks, there also seems to be a misguided attempt on the part of the physicalists to frame the debate between physicalism and dualism, such that dualism is competing with neuroscience or a science whereas in reality, dualism competes with materialism, an opposing philosophical theory (2008: 9). Here Lycan's comment is well taken because primarily the issue that surrounds the debate between dualism and physicalism rarely hangs on what science says about important issues such as those mentioned above (e.g., agency). On the contrary, the heart of the debate between dualism and physicalism is precisely philosophical in nature. But whenever philosophical and scientific domains/issues get conflated, the result often is favoring (ii). But the problem here is that as we saw previously through various examples, (ii) has its own serious limitations.

Moreover, rejecting first philosophy in favor of (ii) also leads us to an ever deeper problem, as I will explain in what follows.

As Moreland persuasively argues, it seems to me that the other contributing factor for the confusion that emerges over the debate between dualism and physicalism has to do with the strong physicalists' failure to make necessary distinctions between four major types of questions. As Moreland puts it, these questions are: (*a*) Ontological questions such as To what is a mental or physical identical? What is a human person? (*b*) Epistemological questions such as How do we come to have knowledge or justified belief about minds and about our own mind? Is there a proper epistemic order to first-person knowledge of one's own mind and third-person knowledge of other minds? (*c*) Semantic questions such as what is meaning? What is a linguistic entity and how is it related to a meaning? Is thought reducible to or a necessary condition for language use? (*d*) Methodological questions such as How should one proceed in analyzing and resolving the first-order issues that constitute the philosophy of mind? What is the proper order between philosophy and science? Should we adopt some form of philosophical naturalism, set aside so-called first philosophy and engage topics in philosophy of mind within a framework of our empirically best-attested theories relevant to those topics? (Moreland, 2008: 158-159).

Jackson, because of his commitment to science (of course, better to say to ''scientism'') ignores to take questions (*a*)-(*c*) at face value. Rather Jackson's SM, tries to merge (*a*)-(*c*) with or reduce them to (*d*). Since (*d*) above helps the physicalists to frame their theory in a way that fits their commitment to the physicalist ontology, questions (*a*)-(*c*) can only be answered in light of the philosophical naturalism that rejects first philosophy. Of course, the suggestion that metaphysical knowledge must be under the umbrella of science echoes very much Quine's well-known doctrine of naturalized epistemology (see e.g., Quine 1969: Ch. 3). Once the physicalist achieves this goal, then appealing to science as stated in (ii) above, will be the only game in town. This is precisely why Jackson's SM hangs on the picture of our world as described by modern science. The question then becomes: what consequence will the physicalist view of science bring to the very position that they advocate?

At least three things come to my mind. First, the physicalists' view of science suffers from scientism, according to which all forms of knowledge have to be empirically based. There is a serious problem here though. That is, the very claim that 'all forms of knowledge have to be empirically based' is not itself a scientific claim, but rather is a philosophical claim about what sort of limit science sets with respect to what counts as knowledge and what does not. So here, there is a clear conflation of a first-order discipline (e.g., science) with the second order philosophical claim about science. Here Lowe's remarks hit home:

> …devotee of scientism…fail to see that science *presupposes* metaphysics and that the role of philosophy is quite as much normative as descriptive—with everything, including science, coming within its critical purview. Scientists inevitably make metaphysical assumptions, whether explicitly or implicitly, in proposing and testing their theories—assumptions which go beyond anything that science itself can legitimate. These assumptions need to be examined critically, whether by scientists themselves or by philosophers—either way, the critical philosophical thinking that must be done cannot look to the methods and objects of empirical science for its model. Empirical science at most tell us what is the case, not what must or may be (but happens not to be) the case (Lowe 1998: 5; see also Moreland, 2008: 158-159).

Insofar as scientism is problematic as stated above, the physicalists can hardly account for everything on the basis of science. Thus, the dualists are within their own epistemic right to advance their metaphysical theory regarding the nature of the self without basing it on science. Second, since emergent properties, agency, etc., are things which cannot be explained (at least primarily) on the basis of science, yet they are part of the fabric of human nature and experience, then the strong physicalist's claim that the accurate picture of our world is the one that only modern science gives us, must be false. So leaving out important things such as agency from the physicalist ontology only shows the inadequacy or the incompleteness of the theory. Third, strong physicalism relies on either a reduction or elimination method to deal with emergent properties. But the problem, as we have already seen, is that such properties are irreducible. Thus the physicalists' methods to either reduce them or eliminate them remain unsuccessful. This is because such methods could do neither job for the physicalist.

But Jackson rejects the above remarks because he thinks that it is a methodological mistake to think that philosophy should revise science. If philosophy revises science, then for

Jackson, this amounts to giving philosophy and metaphysics an immodest role.[129] Jackson's point here is a bit puzzling. He seems to be implying that there is a sort of self-sufficiency for science. If this is what Jackson has in mind, then when he tells us that neither philosophy nor metaphysics ought to revise science, Jackson's claim again amounts to another instance of a philosophical claim about how science should be treated. If I am right, then Jackson's claim should be rejected because it has a self-defeating nature, in that the philosophical claim he makes is incompatible with his own conception of science. To say that it is a methodological mistake for philosophy to revise science is not a scientific statement at all. Hence, if indeed it is a methodological mistake for philosophy to revise science, then all that it implies is that if philosophy were to revise science, then philosophy only ends up violating not science, but scientific naturalism upon which the very physicalism itself is based. I think we need to heed Moreland's remarks, "scientists cannot adequately discuss the central topics in philosophy of mind without making substantive philosophical claims, but philosophers need not discuss scientific data to treat adequately these same philosophical issues," (Moreland, 2008: 173). Of course from such remarks nothing follows to the effect that science is irrelevant in matters of philosophical discussion. But that is another issue. From all this, it follows that Jackson's attempt to unnecessarily compartmentalize science and philosophy remains unpersuasive.

6.5 *Summary*

In this chapter, I discussed Lowe's theory of the self, according to which the self is distinct from the physical body or any part of it, such as the brain. The self is also an intentional agent. Even if the self is a physical thing in virtue of possessing physical characteristics, the self is a bearer of non-physical mental properties. I have also considered objections against Lowe's TS and responded to those objections. I also discussed Jackson's theory of serious metaphysics and argued that Jackson's attempt to locate everything in the physical world is unsuccessful. I have raised objections against Jackson's SM and tried to show why Jackson's rejoinders fail. After having looked at the two models, TS and SM, my own conclusion is that Lowe's TS has a superior explanatory advantage over a range of phenomena discussed in

---

[129] See Jackson (1998: 42-43). Also see, Ludlow, Nagasawa and Stoljar, eds., (2004: 23-24).

this chapter. By contrast, Jackson's SM fails to succeed in what it set out to accomplish, that is, to locate everything in a fundamentally physical world. In light of all this, I would say that there is every good reason to endorse Lowe's TS and reject Jackson's SM. Thus, Lowe's TS is superior to Jackson's SM.

Part IV: The Neo-Aristotelian Account: The Substantial-Self/Person

Chapter Seven

The Neo-Aristotelian Account: The Substantial-Self/Person

7.1 Summary

In the last six chapters, my main emphasis has been on the centrality of substance ontology for *the substantial-self/person question* on the one hand and personal identity on the other. In chapter one, I attempted to establish *the substantial-self/person question*. I also provided a survey of the main issues in the contemporary personal identity debate. In chapter two, I argued, *inter alia*, why substance ontology is central to Locke's theory of personal identity. In chapter three, I argued why despite earning him many followers, Hume's anti-substance ontology project is hardly successful. In chapter four, I argued why the methodology of thought experiments, has paved the way for the radical shift in the subject matter of the contemporary personal identity debate. That is, what I called *the Aristotelian question*, i.e., 'what is X' question has been put in the back seat whereas the puzzle cases and the criterion of identity question have been let to ride in the driver's seat. In chapter five, therefore, based on the observations made in chapter four with respect to the marginalization of the 'what is X' question, I argued that most, if not all, contemporary debates on personal identity have been set off in the wrong direction. In this regard, I have located my discussion within the framework of the currently most influential view of personal identity, namely the relational/neo-Lockean view. I also briefly discussed other contemporary views such as animalism, the constitution view of a person and a substance view of a person in relation to the neo-Lockeanism.

In chapter six, I discussed the ontology of the self in the context of the philosophy of mind. In that chapter, I based my discussion on two representative views, one from the dualist camp and one view from the strict physicalist camp and argued why strict physicalism suffers from a lack of explanatory adequacy with respect to a wide range of issues we grapple with in philosophy of mind (e.g., mental causation, human agency). In the present chapter, my main goal is to pull together the thread of various arguments I have given in the last six chapters in regards to the realist ontology of the self/person, and to sketch out the view of the self/person that I shall seek to defend. This is the view I am calling: *the Neo-*

Aristotelian View of Substantial-Self/Person. As indicated in chapter one in § 1.1.2.3, I will locate this view within the framework of the traditional Aristotelian metaphysics with reference to substance ontology. In doing so, I will attempt to explain what this view amounts to in light of the Aristotelian metaphysics. In chapter eight (which is the final chapter of this thesis), I will consider a couple of objections against this view. I will then conclude the thesis with some remarks about future areas of research.

7.2 *Traditional Metaphysics in a Nutshell*

In his essay entitled, 'On What Grounds What,' Jonathan Schaffer correctly observed that unlike the currently dominant Quinean view, which reduces the core inquiry of metaphysics to what there is, (e.g., whether properties exist, whether meanings exist, and whether numbers exist), metaphysics in the traditional Aristotelian view, is about what grounds what (in Chalmers, Manley and Wasserman, eds., 2009: 347). As Schaffer argues, metaphysics so conceived does not dwell on asking whether properties, meanings, and numbers exist, which most people could agree that they do exist. Rather the real business of metaphysics taken in its traditional sense [*inter alia*], is to deal with whether or not the things that exist are fundamental (*Ibid*.; see also Kit Fine in Tahko 2012: Ch.1).[130] So what does Aristotle take fundamental existent to be? Here at the top of the list we find, primary *ousia*, which Aristotle takes it to be the central subject matter of his metaphysics. So without taking metaphysics seriously, we cannot have an adequate grasp of Aristotle's substance ontology. In *Metaphysics* Γ, Aristotle describes the central concern of 'first philosophy' as follows:

> There is a discipline which studies that which is qua thing-that-is and those things that hold good of this in its own right. That is not the same as any of what are called the special disciplines. For none of the others examines universally that which is qua thing-that-is, but all select some part of it and study what is coincidental concerning that; as for instance the mathematical disciplines. But since we are seeking origins, i.e. the most extreme causes, it is plain that these are necessarily a particular nature's in its own right. If therefore these origins were also sought by those seeking

---

[130] In his essay, Schaffer also points out that contemporary textbooks often introduce metaphysics in the spirit of the Quine-Carnap debate, declaring Quine as a winner. But why start with the Quine-Carnap debate? [As is well-known, for Carnap, traditional metaphysical claims are meaningless owing to their failing to fit into the verificationist epistemological mold]. But Schaffer asks: Why think that the best understanding of metaphysics is to be found in a debate between a positivist teacher and his post-positivist student, both of whom share explicitly anti-metaphysical sympathies? (*Ibid*.) Schaffer's question here entirely resonates with what I have been emphasizing myself all along in this thesis.

the elements of the things-that-are, the elements too are necessarily of that which is qua thing-that-is, not coincidentally. Hence we also have to find the first causes of that which is qua thing-that-is (*Metaphysics* Γ.1. 1003a21).

Here the gist of Aristotle's point is that the object of the investigation of first philosophy is not as limited as some special sciences, which only tend to focus on their respective area of interest. For example, mathematics focuses on things that are countable and measurable. But for Aristotle, metaphysics is a universal science that studies being qua being (cf. Grossmann 1983: Ch. 1). But what does Aristotle mean by 'being qua being'? As S. Marc Cohen remarks, Aristotle's description of 'the study of being qua being' does not imply as if there is a single subject matter—being qua being—which is under investigation. Instead the phrase 'being qua being' involves three things: (1) a study, (2) a subject matter (being), and (3) a manner in which the subject matter is studied, i.e., qua being (S. Marc Cohen 2008: 2). As Cohen further points out, Aristotle's study does not focus on some recondite subject matter identified as 'being qua being'. Rather it is a study of being. In other words, for Aristotle, first philosophy studies beings, in so far as they are beings (*Ibid*.).

Loux (1998: 3-4) argues that metaphysics considers things as existents and tries to specify the properties they exhibit in so far as they are beings. So the chief goal of metaphysics is not just grasp being itself, but also general features of being such as identity, difference, similarity, and dissimilarity that apply to everything that there is. In light of this, Loux further remarks that central to Aristotle's metaphysics is the description of what Aristotle refers to as categories. These are the most general kinds under which things fall. So the business of a metaphysician is to identity those general or highest kinds. But the task does not end here. A metaphysician is also supposed to specify the features unique to each category.[131] The upshot of engaging in such an activity provides us with a map of the structure of all there is (cf. Lowe 2006). This is to say that metaphysics is the study of the

---

[131] For an excellent survey of various conceptions/systems of categories since Aristotle to contemporary philosophers, see Thomasson (2004/2013). For more details on contemporary approaches to systems of categories, see e.g., Chisholm (1996); Lowe (2006); Hoffman and Rosenkrantz (1994); Grossmann (1983); and Ingvar Johansson (1989). For the view that sees ontological categories as systems imposed on the world as opposed to being reflectors of the way the world is, see Westerhoff (2005).

fundamental structure of reality (cf. Sider 2012). Or as Schaffer notes, "metaphysics is about what grounds what," (in Chalmers, Manley and Wasserman, eds., 2009: 347). So given the Aristotelian approach, a metaphysician's success in the investigation of the nature of reality depends on whether or not his/her approach is a realist one. That is to say that whether or not one accepts the existence of a mind-independent reality (cf. Heil 2012; Yagisawa 2011: 270). Aristotle takes a mind-independent reality as the starting point of his metaphysical theorizing. In Aristotle's view, metaphysics underlies our overall approach in our investigation of the nature of reality. In this regard, Fine is right in saying that "I take Aristotle's primary concern in his metaphysical and physical writings to be with the nature of reality rather than the nature of language. I am rarely tempted, when Aristotle appears to be talking about things, to construe him as saying something about words…," (Fine in Lewis and Bolton, eds., 1996: 83-84).[132]

The gist of Fine's remarks above, is that, Aristotle's primary agenda in doing ontology is not to engage in mere conceptual analysis of the words we use to talk about the things or the objects of our investigation.[133] Rather, the central aim of doing ontology has to do with grasping the nature or essence of things—i.e., what makes something what it is (cf. McGinn 2012). Taken this way, language serves only as a medium through which we construct and communicate our concepts regarding the things we investigate (see e.g., *Cat*. Chaps. 1-4). If we adopt such an Aristotelian approach toward ontology, then it has important implications, *inter alia*, for our conception of the self/person. For example, in this case, one of the key questions that can be asked is this: is the notion of the self/person a theoretical construct or primarily a prephilosophical notion? I will attempt to give my own answer for this question in § 7.4.

7.3 *Skepticism against Traditional Metaphysics*

---

[132]In this case, the Aristotelian approach to the investigation of reality stands in sharp opposition to Dummett-Putnam's anti-realism, according to which there is no access to mind independent reality. See Dummett in Hales (1999: Ch. 4); Putnam in Hales (1999: Ch. 5). For an excellent critical discussion on Dummett-Putnam anti-realism, see Loux (1998: Ch.7). I will have something to say, in due time, against general scepticisms raised toward traditional metaphysics.

[133] For example, for philosophers who seem to take conceptual analysis as the main method in doing ontology, see Jackson (1998) and Chalmers (2012).

The hitherto brief description of the traditional Aristotelian metaphysics, however, receives an immediate resistance if one follows Kant's lead. In his *Critique of Pure Reason* Kant opens his discussion by telling us that human reason has paramount relevance in our attempt to understand some fundamental questions about reality, yet at the same time, human reason has its own severe limitations in terms of grasping the nature of reality that goes beyond the boundary of experience (A vii: 5ff, trans., Pluhar 1996). Kant takes traditional metaphysics as a purely speculative discipline which cannot give us access to mind-independent reality (*Ibid*. Bxv: 20). In modern times, it was also rejected by post Hume-Kant influential thinkers such as Quine.

Most notably, a serious attack was launched against the traditional metaphysics by the 1930s and 40s Vienna Circle logical positivists, such as Carnap and others (see e.g., Ayer 1936: 44-47 and Ch. 3). As is well known, logical positivists promoted a movement that restricts the source of genuine knowledge to what can only be empirically verifiable (see e.g., Uebel 2012). Though, at present, the 'verification principle' largely has fallen out of favor, it is safe to say that it is making a comeback (see e.g., Ladyman and Ross, 2007; also see Ladyman, Ross and Kincaid, eds., 2013). Recently, Price in his essay entitled, "Metaphysics after Carnap: The Ghost Who Walks?" argued that metaphysics [traditional] is 'as dead, or at least deflated, as Carnap left it' (in Chalmers, Manley and Wasserman 2009: 322; see also Chalmers 2012). Along similar lines Hofweber also remarks, "The questions that metaphysics tries to answer have long been answered in other parts of inquiry….What metaphysics tries to do has been or will be done by the sciences. There is nothing left to do for philosophy," (*Ibid*. 260). Price's as well as Hofweber's negative attitude towards the traditional metaphysics is also strongly embraced and echoed by Ladyman and Ross in their works (see the references above).

However, Price's and Hofweber's remarks are not only very strong but they are also very questionable in light of the work of notable contemporary neo-Aristotelian advocates of traditional metaphysics (see e.g., Lowe 1998, 2002, 2006; Oderberg 2007; Hoffman and Rosenkrantz 1997; Schaffer in Chalmers, Manley and Wasserman, eds., 2009: chap. 12; Gibb, Lowe and Ingthorsson, eds., 2013; Loux 1998; Heil 2012; Tahko, ed., 2012; Novotný &

Novák, eds., 2014). One can confidently say that Aristotelian metaphysics is back on the stage, contra Price's and Hofweber's claim and all those who are sympathetic to their remarks. But I do not intend to argue for this broad claim, since here my interest is to put the Aristotelian metaphysics to a good use in defending my own account of the self/person. That said, despite some knee jerk reactions of some modern philosophers, traditional metaphysics occupies a central place in our discussion of substance ontology, which provides us with a framework, *inter alia*, for our concept of individuation, identity, persistence, agency and causation (see e.g., Ruggaldier in Corradini, Galvan and Lowe, 2006: Ch.3; also see Lowe 1998: Chs. 4-7).

7.4 *Approach*

In this section, I will explain what is unique regarding *the Neo-Aristotelian View of Substantial-Self/Person*. As indicated in § 7.1., this is the view that gets its inspiration from the traditional Aristotelian substance ontology. But in the mainstream contemporary personal identity literature, Aristotelian substance ontology has been severely marginalized. Even when the neo-Aristotelian personal identity theorists engage with Aristotle's views in the *Categories*, their discussion usually does not go beyond the notion of substance *simpliciter* (see e.g., Wiggins 1980/2001; Lowe 1998; cf. Loux 1998: 123-135). Although as we saw in chapter six, Lowe identifies the self/person with substance, he does not make an explicit connection with Aristotle's *Categories*. Similarly, Wiggins does not attempt to make such a link with the *Categories*, although he argues that the notion of a person coincides with the notion of human being (see e.g., in Perry 1975; see also in Peacocke and Gillett 1987: 4). Furthermore, defenders of animalism, despite talking about the animality of human beings (Aristotle's insight), do not engage with Aristotle's works (see e.g., Olson 1997 and 2007; see also Van Inwagen 1990; cf. Merricks 2001).[134] To my knowledge, I have not come across any work in the contemporary mainstream personal identity literature (see e.g., chapter five footnote # 88), where a deliberate attempt has been made in linking the

---

[134] In personal conversation, Olson told me that he never thought about how his animalism is related to Aristotle's metaphysics. That said, however, Aristotle's views of man, psyche/soul, matter and form and the like are often located in the context of philosophy of mind, where the focus is on appropriating these issues to defend a certain view of mind and body dualism (see e.g., Oderberg 2007: Ch. 9).

notion of the self/person to Aristotle's *Categories*. So in this regard, my *Neo-Aristotelian View of the Substantial-Self/Person* will bridge this gap in the literature, thereby also urging Aristotle's place to be reconsidered in the contemporary personal identity discussion.

Moreover, Aristotle does not discuss personhood in his *Categories*, which is also true of his *Metaphysics*.[135] This is another incentive for me to explore new territory in terms of connecting the notion of the self/person to Aristotle's conception of substance. My primary source in developing my *Neo-Aristotelian View of the Self/Person* will be the *Categories*. Aristotle also discusses the notion of substance in his *Metaphysics*, where he links it to the notion of form and matter, *inter alia*. In § 7.6, I will make some brief remarks about Aristotle's view of form and matter without getting into Aristotle's extremely intricate analysis.[136] That said, I need not also necessarily assume that Aristotle is right in everything he says about substance. Nor do I also assume that Aristotle himself necessarily would endorse everything I say with respect to the connection I make between the notion of the self/person and his theory of substance. For present purposes, I am only interested in the insights that can be gleaned from the above two texts, which I can then put to a good use in developing and defending my own *Neo-Aristotelian View of the Substantial Self/Person.*

As one may recall, in chapter one § 1.1, I identified the notion of the self with the Aristotelian notion of substance. I then subsequently worked my way through Descartes's and Boethuis's notion of the entityhoodness of the self/person. In §§ 7.6 and 7.7, I will sketch out in some detail how I understand the entityhoodness of the self/person.

7.5 *The Starting Point*

As David Lewis (1986: 192-193) once characterized the notion of identity as utterly simple

---

[135] Aristotle's *De Anima* might be the closest of Aristotle's work that can be associated with contemporary discussions on personhood (see Michael Durrant 1993; Ronald Polansky 2007).

[136] For exegetical issues and controversies on Aristotle's view of substance, see, e.g., Gill (1989); Terence Irwin (1988); Lewis and Bolton, eds., (1996); Allan Code (1978); Loux (1991, 2008); Theodore Scaltase (1994, 2010); Stephen Makin (2006); and Ross, vol. 1 & II (1924).

and unproblematic, I too want to claim that the notion of the self/person is utterly simple and unproblematic. However, my claim here immediately faces resistance from those who think that the notion of the self is deeply perplexing. For example, Rudd remarks:

> One might think, naively, that the concept of the self is not one that should present great problems for us. After all, we are all selves, so shouldn't we all just know what it is we are? But the concept of selfhood is in fact among the most troubling and contentious in philosophy. And not only in philosophy—the concept has been puzzled over by neurologists, psychologists, psychiatrists, criminologists, sociologists, theologians, historians, and—in rather different ways—by novelists, dramatists, and film-makers (2012: 1).

Here I can think of two senses in which one may or may not agree with Rudd's remarks. The sense in which one may agree with Rudd has to do with the obvious point that the notion of the self is understood differently in different contexts (see also e.g., chapter one footnote # 2). So in such diverse contexts, accounts of the notion of the self could be mutually exclusive and incompatible. Hence, such accounts could be perplexing, indeed! But what does it follow from such diverse accounts of the notion of the self? For example, does it follow that we do not have a pre-philosophical sense of our own selfhood? If Rudd's remarks in the above passage are taken to give an affirmative answer to this question, then one (which includes myself) is within his/her own epistemic right to disregard Rudd's remarks. This is because no matter how much perplexing and controversial the notion of the self could be, as we shall see, selfhood/personhood is always presupposed at the background. In this case, the reality of the unavoidable pre-philosophical sense of oneself is beautifully captured by Taylor. Here is a passage from his *Metaphysics*:

> However unsure I may be of the nature of myself and of the relation of myself to my body, I can hardly doubt the reality of either. Whether I am identical with my body, or whether I am a spirit, I cannot doubt my own being, cannot doubt that I am part of the world, even prior to any philosophical reflection on the matter. For surely if I know anything at all, as presumably I do, then I know that I exist. There seems to be nothing I could possibly know any better. And this is, of course, quite consistent with my great ignorance as to the nature of that self of whose existence I feel so assured. I know, further, that I have a body. I may have learned this from experience, in the same way that I have learned of the existence of innumerable other things, or I may not have; it is, in any case, something I surely know. I may also have only the vaguest conception, or even a totally erroneous one, of the relationship between myself and my body; I can nevertheless no more doubt the reality of the one than the other. I may also be, as I surely am, quite ignorant of the nature and workings of my body

and even of many of its parts, but no such ignorance raises the slightest doubt of its reality (1974: 11-12).

Taylor's remarks here hinge on one basic premise, that is, we all have a pre-philosophical notion of a sense of selfhood/personhood (see also, Moore 1959; Elder 2004; Hoffman and Rosenkrantz 1997: 1-7; Paul 1961: 19; O'Hear 1985: 216-217). Such pre-philosophical sense of one's own selfhood/personhood is primitive in that we do not arrive at it on the basis of analysing other concepts. Rather such sense of our own selfhood/personhood is our starting point to make sense of other things. Put differently, it is also what allows us to distinguish ourselves from other people on the one hand, and other things around us on the other. If this is right, then a pre-philosophical sense of our own selfhood/personhood is an important contributor to self-individuation. It seems then that a pre-philosophical notion of one's selfhood/personhood can be taken as a base-level or primary epistemic source for other aspects of diverse self-conceptions we may see in different contexts. In this case, Rudd's remarks we saw earlier can be said to make sense only if they are grounded in the base-level epistemic source. Here by 'base-level epistemic source,' I mean to refer to what I said earlier that our pre-philosophical sense of our own selfhood/personhood is *primitive*. It is such primitiveness that serves as a base-level epistemic source for our conception of the selfhood/personhood.

If we grant such primitiveness to play the primary role towards our pre-philosophical conception of the self, then we should be prepared to reject any suggestion that our basic notion of selfhood or personhood is the result of some sort of theoretical achievement or theory laden (see e.g., Armstrong 1980: 64-65; Churchland 1988: Chaps. 2-4; cf. Harré in Kolak and Martin 1991: 370). I will give reasons why there is a sense in which a theory-laden conception of our basic sense of self is deeply mistaken. I will return to this issue in § 7.6. In saying this, however, I am not implying that a theory laden understanding of one's own self is of no value. Quite the opposite. Rather my claim is that, a theory laden conception of one's sense of self is always *posteriori* to that of a pre-philosophical conception. It is difficult to see how one can even come up with any credible story about one's own basic sense of self in abstraction from a pre-philosophical conception. In this regard, Aristotle's approach

241

is instructive, since as Hoffman and Rosenkrantz noted, the entities Aristotle lists in his categories all belong to commonsense or folk ontology (1997: 1). It is in light of such considerations, that I too want to sketch out my *Neo-Aristotelian View of the Substantial-Self/Person*. Before I do that, I will discuss Aristotle's theory of substance in his *Categories*.

7.6 *Substance in the Categories*

As we saw in chapter one § 1.1, the Greek word *ousia*, the noun from the verb 'to be' (*einai*) has a technical meaning, which its equivalent English term 'substance' fails to capture. This is because the English term 'substance' could be misleading, implying a kind of *stuff* (Cohen 2009: 197). The closest meaning that does justice to Aristotle's *ousia* is 'being' or 'entity' (J.L. Ackrill 1963: 77). *Ousia* can also be construed as 'reality' or 'fundamental being' (Cohen 2009: 197) or 'an ontologically basic entity' (Loux 1991: 2-3).

Right from the outset the questions remain: what confers 'substance' the status of an ontologically basic entity? What does it mean to say that 'substance' is fundamental or an ontologically basic entity? Why do other non-substantial entities (e.g., qualities) lack similar status, i.e., an ontologically basic entity? For Aristotle, these and similar other questions are critical in his characterization of *ousia*. In his *Categories*, Aristotle mentions *ousia* as one of the ten most general kinds (categories) under which entities in the world fall. As Aristotle states:

> Of the things said without any combination, each signifies either substance or quantity or qualification or relative or where or when or being-in-a-position or having or doing or being-affected. To give a rough idea, examples of substance are man, horse; of quantity: four-foot, five-foot; of qualification: white, grammatical; of a relative: double, half, larger; of where: in the Lyceum, in the market-place; of when: yesterday, last-year; of being-in-a-position: is-lying, is-sitting; of having: has-shoes-on, has-armour-on; of doing: cutting, burning; of being-affected: being-cut, being-burned (1963: 1b25).[137]

Before we see whether or not there is anything distinctive about *ousia*, it is crucial to keep in mind that the categories above classify things not words (Ackrill 1963: 73-78). Irwin also

---

[137] Here and in his *Topics* 1.9 103b20 Aristotle lists ten categories. However, in his *Physics* and *Metaphysics* K. 12 1068a8, Aristotle mentions seven categories.

claims that for Aristotle, the items in the categories are beings (*onta*), whose classification are based on their relation to a non-linguistic subject (*Cat.*, 1963:1a20ff) (Irwin 1988: 52-53; also see Ross Vol. 1, 1924: 1XXXII- XC). So even if Aristotle makes distinctions between words, they do not reflect common grammatical distinctions (Irwin 1988: 52-53).[138] This shows that Aristotle's focus is on the study of 'beings' as opposed to the language we use to talk about them. In this regard, the category of substance remains at the center of Aristotle's investigation (see again § 7.2).

In *Categories* 5, Aristotle distinguishes between *primary* and *secondary* substances. An example of primary substance is the individual man or the individual horse; whereas secondary substances are the species and genus, for example, man and animal respectively (1963: 2a11-19). But why should we think that the individual man (or the individual horse) is a primary substance as opposed to a certain quality such as the individual white? For Aristotle, the reason is because unlike *the individual man*, that is neither said of a subject nor in a subject (1963:2a11), *the individual white* is in a subject, but is not said of any subject (see e.g., 1963:1a23-8). Aristotle uses the phrase 'in a subject' technically. As he states: "By 'in a subject' I mean what is in something, not as a part, and cannot exist separately from what it is in" (1963: 1a24-25).

Scholars still debate about Aristotle's use of the two phrases, i.e., 'something said of a subject' and 'something in a subject' (see e.g., Hoffman and Rosenkrantz 1997: 14-16; Ackrill 1963: 74-76; Ross v.1 [1924], 1xxxii-xc). But as Ackrill rightly remarks, in Aristotle's view, the phrase 'said of a subject' distinguishes species and genera from individuals (*Ibid.*). On the other hand, the phrase 'in a subject' serves to distinguish qualities, quantities, and items in other dependent categories from substances, which do not depend on other things for their existence (*Ibid.*). The dependency relation that exists between primary substances

---

[138] *Ibid*. Grammatically similar words may have different roles to play in terms of signifying items in different categories. For example, as Irwin remarks, for Aristotle, if we take two grammatically similar words 'man' and 'musician', the former signifies an item only in the category of substance (e.g., a human being) whereas the latter signifies an item in two distinct categories, i.e., substance, a human being as well as one of his qualities, musicality. Aristotle's conception of categories stands in sharp contrast to that of Kant's, since for Kant categories are only features of our thought (see e.g., Carr 1987: 2-7).

and items in other categories is ontological, in that if *X* inheres in a subject *Y* then, *X* cannot exist independent of *Y*.

In this case, for Aristotle, only substances enjoy the privilege of being 'not in a subject.' That is to say that "all the other things are either said of the primary substances as subjects or in them as subjects," (1963: 2ba34-35). To use Aristotle's own example, the animal is predicated of man as well as the individual man. Similarly, colour is in body as well as in an individual body. For Aristotle, the upshot of such predication relation between substances and other things predicated of substances boils down to this: "if the primary substances did not exist it would be impossible for any of the other things to exist," (1963: 2a43-45). Hence, primary substances are ontologically basic in that they do not require other things for their existence whereas the converse is not true.

In light of this, what can we say about the nature of the relation that holds between the primary substances and the secondary substances? Before we see Aristotle's response to this question, here we can take (even if we don't come across the terms in the *Categories*), primary substances as particulars, whereas secondary substances as universals. This is consistent with Aristotle's definitions of such terms in *De Interpretatione*: "Now of actual things some are universal, others particular (I call universal that which is by its nature predicated of a number of things, particular that which is not; a man, for instance, is a universal, Callias a particular)," (1963: 17a38).

Aristotle claims that if we want to say what the identity of a concrete particular such as Socrates is, then the appropriate thing to say is that Socrates belongs to a kind *man*. On the other hand, it is wrong to take properties such as 'white' or 'runs' or anything like that to constitute the identity of Socrates (1963: 2b29ff). This is because to be a *human being* for Socrates is an essential property without which Socrates would lose his existence. But the same cannot be said regarding properties such as being *white* or being *a runner*. For example, Socrates may lose the property of being *a runner* without thereby ceasing to exist or losing his identity. This is because, being *a runner* is only an accidental property that Socrates has or instantiates. An accidental property is a property, which a subject can lose

without losing its existence. But if we take away from Socrates a kind *man* (which is an essential property), then Socrates no longer retains his identity much less his existence.

Aristotle argues that secondary substances, are not in a subject, in this case in a primary substance in the manner entities in other categories are (e.g. qualities). As he states, "It is obvious at once that they [secondary substances] are not in a subject. For man is said of the individual man as subject but is not in a subject: man is not *in* the individual man," (1963:3a11-12).[139] Even if secondary substances are not said to be in a subject, ontologically they still depend on primary substances. Aristotle tells us that the manner of such dependency relation is not that of the inherence. As he puts it:

> Every substance seems to signify a certain 'this'. As regards the primary substances, it is indisputably true that each of them signifies a certain 'this'; for the thing revealed is individual and numerically one. But as regards to the secondary substances, though it appears from the form of the name—that a secondary substance likewise signifies a certain 'this', this is not really true; rather, it signifies a certain qualification, for the subject is not, as the primary substance is, but man and animal are said of 'many things"(1963:3b10-19).

Here Aristotle anticipates that there may be a temptation to lump secondary substances with the category of quality, since secondary substances are also labeled as 'qualification.' But Aristotle rejects such a move on the basis of the difference that exists between the category of quality and that of the category of substance. As he states:

> However, it [secondary substance] does not signify a certain qualification, as white does. White signifies nothing but a qualification, whereas the species and the genus mark off the qualification of substance—they signify substance of a certain qualification (1963: 20-24).

Commenting on this passage, Cohen remarks that Aristotle seems to be saying that what makes species and genera secondary is that they are kinds or collections. A collection of individuals stands for species, whereas a wider collection of the individual members of the species that fall under it is the genus. As Cohen further explains, for Aristotle, without those

---

[139] Similarly, Aristotle extends the same conclusion to *differentiae* of a substance— "roughly, the properties that are in the definition of the substance," (Cohen 2009: 198). As he states: the differentia also is not in a subject. For footed and two-footed are said of man as subject but are not in a subject; neither two-footed nor footed is *in* man.

individuals, there would be no species, and without species there would be no genera (Cohen 2009: 199). This means that for the species 'man' to exist, there must be the individual men. According to Aristotle, then real things are the individual men. What then are the species and genera? For Aristotle, the species and genera are in the words of Cohen "simply the way specimens are classified and organized," (*Ibid*.). At this point one may ask whether or not there is a mutual ontological dependence between them. Cohen claims that although Aristotle does not comment on this point in the *Categories*, he explicitly talks about it in the *Topics*, "It is impossible for a thing still to remain the same if it is entirely transferred out of its species, just as the same animal could not at one time be, and at another not be, a man," (IV.5 125b37-40)".[140] As Cohen notes, "the fact that the *said-of* relation seems to amount to what Aristotle elsewhere calls essential predication makes this idea even more plausible," (2009: 199). Here Cohen is referring to Aristotle's point in the *Categories*—"if something is said of a subject both its name and its definition are necessarily predicated of the subject," (2a19-20)." For Aristotle, telling the definition of X amounts to explaining the essence of X (*Topics*, 1.5 101b38, VII.5 154a31 as quoted in Cohen (2009: 200)).

Aristotle also claims that there is nothing contrary to substances. As he states, "there is nothing contrary to an individual man, nor yet is there anything contrary to man or to animal," (1963:3b24-28). Aristotle further claims, "this is not peculiar to substance but holds of many other things also, for example, of quantity [such as]…four-foot or ten" (1963:3b29-31). Aristotle also claims that substances do not come in degrees. However, this does not mean that one thing is not more substance than the other. For example, Aristotle takes a species, *man* to be more substance than the genus, *animal*. This is because, for Aristotle, a species *man* is much closer to the individual man than *animal* (see further 1963: 2b7ff). In talking about why any substance does not admit of degrees, Aristotle remarks:

> For one man is not more a man than another, as one pale thing is more pale than another and one beautiful thing more beautiful than another. Again, a thing is called more, or less, such-and-such than itself; for example, the body that is pale is called

---

[140] As quoted in Cohen (2009: 199).

more pale now than before, and the one that is hot is called more, or less, hot. Substance, however, is not spoken of thus. For a man is not called more a man now than before, nor is anything else that is a substance. Thus substance does not admit of a more and a less," (1963:3b30-48).

The other distinctive feature of substance has to do with numerically one and the same substance being able to receive contraries. Aristotle claims that in no other case can we think of anything that is numerically one and the same and yet receives contraries. For example, for Aristotle, "colour which is numerically one and the same will not be black and white, nor will numerically one and the same action be bad and good; and similarly with everything else that is not substance,"(1963:4a10-17). By contrast, Aristotle claims, "an individual man—one and the same—becomes pale at one time and dark at another and hot and cold, bad and good. Nothing like this is to be seen in other cases," (1963:4a19-21). Aristotle's point here is that substance remains numerically the same through qualitative change.

In the *Categories*, Aristotle does not treat primary substances as complex bodies. But he does treat primary substances as complex bodies in his *Metaphysics* (see further Gill, 1989: Ch.1). Because of such a shift in Aristotle's thinking, there is a debate on whether complex entities, i.e. entities with combination of matter and substantial form are primary substances. Yet Aristotle's own commitment to the primacy of substance over stuff still stands (see e.g., Loux 1991). Since for Aristotle, stuff has only potentialities as opposed to actualities, stuff fails to be basic. To say that rock is stuff of a building is to say that rock has only a potential to become a building. By contrast, actuality is prior to potentiality (Book Θ 8, 1049b 18-25). Since substance is actuality, it follows that substance is prior to stuff. That said, however, for Aristotle, matter is what *something is made of* or as some call it '*proximate matter*' as opposed to being a particular kind of space filling *stuff* (Lowe in Oderberg 1999:2-3)*.* Moreover, Aristotle does not subscribe to the notion of the ultimate constituents of bodies, such as atoms, since he does not endorse atomism. As Robinson puts it, 'matter' for Aristotle is "the name for whatever, for a given kind of object, meets a certain role or function, namely that of being that from which the object is constituted," (2009). For example, if we take relative to the human body, matter is not the

molecules/atoms that make up the human body but rather flesh and blood. Similarly, the matter of the table is the wood from which it is. On the other hand, the 'form' answers the question, 'what kind of thing a certain object is' (Robinson 2009). More precisely, as Cohen remarks that the 'form' Aristotle has in mind is not to be identified with a mere property a substance happens to bear. Rather, for Aristotle, the 'form' is constitutive of what the *ousia* by its very nature is (Cohen 2009; see also *Met*. 3 1028b37-1029a6).

7.7 Sketching Out the Neo-Aristotelian View of the Substantial-Self/Person

As I explained in chapter one §§ 1.1.2.3 and 1.1.2.4, my defence of the *Neo-Aristotelian View of the Substantial-Self/Person* will be constrained by two conditional statements, namely: (CS-1) *If the Aristotelian substance ontology is true, then….*; and (CS-2) *If I am a self/person, then I am a certain kind of entity, i.e., a substantial entity.* Notice that (CS-1) and (CS-2) are related in that the former provides a framework within which the latter will be developed. Taken this way, the *Neo-Aristotelian View of the Substantial-Self/Person* attempts to spell out the entityhoodness of the self/person in terms of the category of substance. As we may recall, in chapter one (see e.g., §§ 1.1, 1.1.2.3 and 1.1.2.4), I explored four ways through which one may attempt to give an analysis of the notion of the self in terms of the category of substance. The four ways have to do with different readings of the copula 'is' namely: (a) the 'is' of predication; (b) the 'is' of identity; (c) the 'is' of instantiation; and the 'is' of constitution. Having examined (a)-(d) in light of Aristotle's conception of substance as explained in the *Categories*, I argued why the readings of (a), (c), and (d) would fail as an answer to *the substantial-self/person question.* I then argued why alternatively, (b) can be a good answer to *the substantial-self/person question*. Given (b), I argued that the 'is' in the statement, 'the self is a substance' must be understood as the 'is' of strict numerical identity.

A. *The Category of Substantial Entity*

To say that *X* is a substantial entity is to say that, *inter alia*, *X* belongs to a distinct kind of an ontological category, namely the category of substance. The question remains: what follows from X's belonging to the category of substance? Given Aristotle's account of substance as discussed in § 7.6, a number of things seem to follow:

(1A) X is an ontologically basic entity or fundamental being.

(2B) X stands in certain relation to nonsubstantial entities that exist in other categories.

(3C) X persists both at a time, as well as over time, through qualitative change as numerically one and the same entity.

(4D) X does not come in degrees.

(5E) X is a frame of reference to its own intentional acts—whether those acts are being done in thought, verbally, or otherwise.

(6F) Talk about the category of X is not simply talk about the terms *simpliciter* involving X. Rather it is talk about the very *X* itself as communicated by those terms.

(7G) There is a sense in which X is its own principle of individuation.

In light of the affirmative answer given in chapter one to *the substantial-self/person question*, the above (1A)-(7G) together apply to the notion of the self/person. In this case, we can substitute the variable 'X' in (1A)-(7G) for the term, 'the self/person'. This gives us the following modified desiderata:

(1A') The self/person is an ontologically basic entity or fundamental being.

(2B') The self/person stands in certain relation to non-substantial entities that exist in other categories.

(3C') The self/person persists both at a time, as well as over time, through qualitative change.

(4D') The self/person does not come in degrees.

(5E') The self/person is a frame of reference to its own intentional acts—whether those acts are being done in thought, verbally, or otherwise.

(6F') Talk about the category of the self/person is not simply talk about the terms *simpliciter* involving the self/person. Rather it is talk about the very *X* itself as communicated by those terms.

(7G') There is a sense in which the self/person is its own principle of individuation.

As we shall see, the above (1A')-(7G') are interrelated in a sense that what we say about one will have direct bearing on what we say about the other. Moreover, (1A')-(7G') collectively show the centrality of the category of the self/person. According to the *Neo-Aristotelian View of the Substantial Self/Person*, such centrality of the self/person, *inter alia*, consists in the self/person being the subject of experience, the unifier of experiences and a causal agent (Cf. White 1991).

B. The Analysis of the Desiderata of the Neo-Aristotelian View of Substantial Self/Person

Given (1A'), the self/person is an irreducibly basic substantial entity. I understand the irreducibility of the self/person from three standpoints: (i) the self/person is our starting point of our inquiry into the nature of other things. In saying that the 'self/person is our starting point,' I mean that the reality of the self/person is something that is taken for granted. In other words, the reality of the self is not a theoretical construct. Rather, the datum for the reality of the self/person is already given in common-sense ontology. As Hoffman and Rosenkrantz nicely put it:

> Our culture possesses a single ordinary, commonsense, or "folk" conceptual scheme which has certain ontological presuppositions. Foremost among these presuppositions is the idea that there are enduring things, or individual substances: continuants such as human persons….The idea that there are such substantial beings is at the core of this commonsense or folk ontology (1997: 1).

In light of this, Hoffman and Rosenkrantz further remark:

> Any ontologist must begin as a point of reference with a consideration of folk ontology, even if in the end he or she revises it in some way. If entities of a certain kind belong to folk ontology, then there is a *prima facie* presumption in favor of their reality. Since living and nonliving things or individual substances are a part of folk ontology, there is a presumption in favor of their existence. Belief in the existence of such entities is justified so long as this presumption is not undermined. Thus, those who deny their existence assume the burden of proof (1997: 7).

(ii) If we grant the suggestion that the reality of the self/person is the starting point of our inquiry into the nature of other things, then the self/person has an ontological primacy over other things existing in other categories. As we recall, this is a central notion in Aristotle's *Categories* (see § 7.6). Given the Aristotelian approach, for example, an entity *E*, has an ontological primacy over other entities *Ys*, just in case *Ys* are dependent on *E* in some ontologically important sense. To see this point, we need to explain in what sense as stated above in (2B'), the self/person stands in relation to non-substantial entities that exist in other categories. Quite generally, entities that could not exist without being conjoined with other objects cannot be identified with an individual substance. Such things are often referred to as dependent particulars. As Hestevold nicely sums up:

> There do exist edges, scratches, shorelines, waves, shadows, lines of latitude, holes and their boundaries, reflections, wrinkles, and smiles. Such things, however,

apparently *depend* on other things such as marble cubes, tables, bodies of water, ships, doughnuts, tires, mirrors and faces. Though a cube can obviously exist detached from all other three dimensional objects. A body of ocean water could exist without a wave, but no ocean wave could exist without the water. A tire can exist without its shadow, but the shadow cannot exist without the tire nor can the shadow exist without some other object that shadows it. The grandmother's reflection depends, in some sense, both on a reflective surface and on the face reflected; and the grandmother's smile and wrinkles depend on her face, which existed years ago independently of that smile and those wrinkles (Hestevold in Hales 1999: 415; see also Hoffman and Rosenkrantz 1994: Ch. 1).

The gist of Hestevold's observations in the above passage has to do with a certain dependence relation that obtains between substantial and non-substantial entities. In this case, the dependence relation is existential, in a sense that non-substantial entities cannot exist unless substantial entities sustain their existence. Taken this way, the existential dependence relation that holds between non-substantial entities and substantial entities is asymmetrical, as illustrated in the above passage (see also Fine 1995: 269-290; Simons in Oderberg 1999: 23). In a similar way, the dependence relation that exists between the substantial self/person and the non-substantial mental particulars, such as beliefs, desires, intentional states, phenomenal consciousness, and the like is existential in nature. Contra Hume, it is not possible for mental states, such as perceptions or other mental states to exist without an owner. Anyone who thinks otherwise is setting himself/herself up to embracing most likely an indefensible position. For example, if ownerless mental states exist, then how are we supposed to individuate them?

These are just some of intuitively compelling points and questions that can be raised to raise suspicion about the plausibility of the very idea of ownerless mental states. Of course, defenders of a bundle view can claim that various ordinary particulars are constituted by properties that are associated with them (Loux 1998: Ch. 3). As we saw in chapter two, the advantage often attributed to a bundle view is far from clear compared to that of the substance based view of ordinary particulars (see e.g., Hoffman and Rosenkrantz 1994: Ch. 1; Loux 1998: Ch. 3; cf. Cleve in Hales 1999: Ch. 26). Furthermore, if we follow Aristotle's lead, we should reject the bundle view for one key reason. That is, although a substance has features, those features are not part of a substance. For example, the legs of a chair are

certainly parts of a chair. However, the size and shape of a chair are not parts of a chair. If this is right, then bundle theorists are making a category mistake in claiming that a certain concrete object like a substance is constituted by its properties. In light of such considerations, what confers the self/person ontological primacy over mental states has to do with the latter's dependence on the former for their existence.

(iii) Given the Aristotelian approach, individual concrete entities that belong to the category of substance are primitive in that they are not further analysable in terms of more fundamental things than themselves. Similarly, since the self/person is a concrete substantial entity, it too is primitive. For example, in his *Individuals*, Strawson characterizes the concept of a person as a type of entity to which both predicates that ascribe states of consciousness on the one hand, and predicates that ascribe corporeal/physical characteristics, are equally applicable to a single individual of that type (1959: 101-102). Strawson calls the former P-predicates and the latter, M-predicates. An example of the P-predicates includes, 'is in pain', 'is going for a walk', 'is thinking hard', etc. An example of the M-predicate includes, 'weighs 10 stone,' etc., (*Ibid*. 104). Strawson thinks that despite the fact that we have two distinct categories of property (i.e., P-predicate and the M-predicate), we do not have two distinct subjects as their bearers (see e.g., *Ibid.* 98-101). Strawson's basic point is that even though the properties that are attributed to a person are analysable in terms of a person, the converse does not hold. It is in this sense that Strawson's notion of the primitiveness of the concept of a person needs to be understood. Although I embrace Strawson's view about the primitiveness of the notion of a person, I sharply disagree with his proposal about only one subject being a bearer of two distinct properties. As we recall, in chapter six, I defended the notion that the self/person is distinct from the body that embodies it or any part of it such as the brain. In light of this, I also argued why it makes sense to think that the bearer of physical properties is physical substance, whereas the bearer of mental properties is mental substance. If the reasons we saw in chapter six in defence of these notions are right, then Strawson's proposal of a single subject as being a bearer of two distinct properties turns out to be unconvincing.

Given (3C'), the self/person persists by being wholly present at different times. But what is the basis for such a claim? Loux claims that there are analogies between an Aristotelian substance theory and an endurantist account of temporal persistence. Even if endurantists do not necessarily embrace an Aristotelian substance theory, as Loux points out, the two theories share a lot in common. The main reason for such commonality between these two theories has to do with the fact that both are rooted in our pre-philosophical experience of the world. Loux claims that our pre-philosophical conception of the world is one that leads us to come up with various categories, namely trees, cats, human beings and so on. Given our pre-philosophical conception of the world, we take such ordinary familiar objects as fully real, as opposed to them being things that are constructed out of things that are more real. Furthermore, Loux claims that our belief in the ontological irreducibility of ordinary familiar objects and our belief in their literal diachronic identity is presupposed by our pre-philosophical conception of the world (Loux 1998: 230-231). The self/person is a substantial entity that persists over time through qualitative change.

However, a person does not persist by having temporal parts of any kind, for example, as Lewis argues (1986); or by having stages, for example, as Hawley argues (2001); or by having instantaneous stages as Sider argues (2001). In the Aristotelian picture, a person/self persists through time by being wholly present at different times. Since persons are concrete persisting substantial entities, they cannot have temporal parts like processes or events. Although a person's career can be said to have temporal parts, persons themselves do not extend in time and thus have no temporal parts. For example, as Olson remarks, "your *career* or history may be stretched out in time, and consist of earlier and later parts, a first half and a second half; but you yourself are not," (1997: 5). Along similar lines Shoemaker also argues that, when a particular thing persists over an interval of time, there is a series of events and property instances that constitutes the career of the particular thing during that interval. In this case, the events in the series are understood as causally connected series of property instances. But as Shoemaker further argues, persisting things should not be identified with their careers. In saying this, Shoemaker is directly objecting to the perdurantists' views of persistence (Shoemaker in Gasser and Stefan 2012: 124; see also Olson 2007: Ch. 5).

253

If this is right which I believe that it is, then it leads us to (4D'), which states that the self/person does not come in degrees. What this means is that, *inter alia*, that no self/person is more person/self than the other. A self/person is what it is only in virtue of being the kind of entity that it is, in this case, a substantial entity. In chapter eight, I will spell out in detail the implications of this claim in my response to three contemporary philosophers who defend a degreed notion of a person/self. According to these philosophers, not all human beings are persons. This is because, as these philosophers argue, to be included in the class of persons, certain conditions have to be met. That said, in saying that the self/person does not come in degrees, it is also assumed that the self/person has a determinate identity, in the strict literal sense of the term. Here my view echoes that of both Butler's and Reid's, who argued as we saw in chapter five that personal identity is determinate.

As we saw in chapter five, indeterminacy in personal identity is entertained in the context of certain types of puzzle cases. But in the Aristotelian approach, puzzle cases such as the fission of selves/persons look highly counterintuitive. The Aristotelian argues that persons/selves are not the kinds of entities that undergo fission. There is no evidence that shows that persons naturally undergo the biological process of fission like an amoeba. So no matter how attractive entertaining the possibility of persons as entities that could undergo fission may be, the Aristotelian remains unmotivated to draw any metaphysical conclusion about persons based on such scenarios. The Aristotelian is also within her own epistemic right not to even use bogus science fiction scenarios in developing her ontology of the self/person. In this case, the Aristotelian can say that there should not be any prerequisite, philosophically speaking, to use puzzle cases as a methodology proper in thinking about the ontology of the self/person. But this does not mean that the Aristotelian does not engage with those who develop their ontology of the self/person based on considerations of various sorts of puzzle cases. Nor does it also mean that, if she wishes, the Aristotelian cannot use puzzle case scenarios insofar as they are utilized within the proper constraint as suggested in chapter four. The point here is that given (4D'), personhood is not something that someone has and loses depending on whether or not that person meets certain conditions; rather personhood is something a person has by belonging to a kind human.

Again that does not mean that a person is not distinct from the animal body that embodies it.

Given (5E'), the self/person is a frame of reference to its own intentional acts—whether those acts are being done in thought, verbally, or otherwise. To get this point, first we need to remove certain conceptual obstacles from our way. As we recall, in chapter one, I rejected using the third personal language in talking about the self/person. By this, I do not mean that we cannot use third personal language when we talk about other people as in for example, 'my friend is depressed'. My rejection of the third personal language in talking about the self has to do with when I myself talk about myself using the term 'self,' in third personal language. I rejected this way of talking about the self, because it is misleading in giving a false impression that the self is something to be had as some sort of commodity.

For example, suppose that I am a Cartesian soul. Do I have a soul or am I a soul? Most people commonly say that they have souls or selves. Such way of talking, if it is taken as a manner of speech, may not be problematic. But if by 'having a soul' one means literally he/she has a soul, then such talk must be wrong. This is because, if one is a Cartesian soul, then one is essentially a soul. So one should not talk about soul using the locution of 'having,' unless he/she uses it as a manner of speech. However, one can say that one is an 'ensouled being' thereby implying that one is an embodied being (cf. Moreland and Rae 2000: Part I). What seems to me to be the right thing to say then is this: it makes sense for me to say that I have a body, assuming that I am not essentially my body. But if I am not essentially my body, then I am essentially, say, for example, a Cartesian soul. On the flip side, suppose that I am not a Cartesian soul. In that case, then I am an essentially material being. Or following Olson, I can say that I am essentially an animal (see e.g., Olson 1997; 2007). Again suppose that I am neither a Cartesian soul nor a purely material being, then in that case I am an atomic simple (see e.g., Moreland 2009: 119-120). So there are many other ways by which one could characterize one's identity depending on what one takes himself/herself essentially to be.

The point of raising the above clarificatory points is intended to show that the notion of the self/person, if it is taken from the first person perspective, cannot be eliminated without presupposing the very entity one tries to eliminate.[141] This means that the notion of the self/person is so central to our conception of ourselves, as human beings. Take for example, locutions such as 'self-deception', 'illusory self', 'no self' and so on (cf. Sidertis, Thompson and Zahavi, eds., 2011; Sorabji 2006: Ch. 16). If one makes such assertions from the standpoint of one's own first person perspective, each of these locutions presupposes the reality of the self. One can only engage in self-deception, if one is a kind of entity that is capable of doing such a thing. Similar, if I say 'the self is illusory,' then for my claim to have any hope of making sense, I must assume that it is a sentence that is produced by a thinking, conscious, reflective being and so on. Furthermore, if I utter a sentence, 'there is no self', one can always ask: who is doing the denying? The answer we get for this question is clear, that is, a denier must be an entity that is capable of doing such an act of denying. In light of this, if the notion of the self/person is understood from the standpoint of the first person perspective, it always functions as its own frame of reference. As Kosso remarks in another context:

> When I say the train is moving, it is simply understood that it is moving with respect to the platform. But it is always an incomplete description to say that something is moving or sitting still. Moving in relation to what? Saying simply that the earth is moving is as meaningless as saying that the earth is bigger. Bigger than what? (1998: 35).

Here the gist of Kosso's remark is that notions such as 'motion', 'bigger' necessarily presuppose things that move or things that are bigger. Similarly, any attempt to deny the reality of the self or eliminate it from one's ontology stubbornly presupposes it. If this is right then, the notion of the self/person is so fundamental that trying to explain it away is not as easy as it might seem to be. If my hitherto observations are right, then they further consolidate my earlier claim about the primitiveness of the notion of the self/person. This is one of the main reasons why the Aristotelian thinks, as indicated in (6F') that talk about the

---

[141] In light of this, I reject the view of those philosophers who argue that ordinary familiar things do not exist, including people. These philosophers base their arguments on Sorties Paradox. For such nihilism, see e.g., Unger (1979); Stone (2005); and Heller (1990: Ch. 3).

category of the self/person is not simply talk about the terms *simpliciter* involving the self/person. Rather it is talk about the very self/person itself as communicated by those terms. So for the Aristotelian, the category of substance is not the category of concept. This way of thinking stands in sharp contrast to how the notion of person is used by most contemporary philosophers. As we recall from our discussion in chapters one and five, the functional notion of a person allows the notion of a person to function like an indefinite description (e.g., a President of Ethiopia), thereby constraining its application to whatever that satisfies it (see Daly 2013: 14-15). But the Aristotelian rejects a functional notion of a person. Given that the Aristotelian takes the category of the self/person as a category of being, what is its implication for the individuation of such beings?

The above question brings us to (7G'), which states that there is a sense in which the self/person is its own principle of individuation. Quinton once wrote that individuation is one of the four problems of substance (1973: 4). Can the Aristotelian address the problem Quinton points out? The answer is yes. There are two senses of the term 'individuation', namely a metaphysical and an epistemic sense (Lowe in Loux and Zimmerman, eds., 2003: 75). As Lowe remarks, taken in an epistemic sense, "for someone to *individuate* an object…is for that person to 'single out' that object as a distinct object of perception, thought, or linguistic reference," (*Ibid*.). On the other hand, taken in the metaphysical sense, as Lowe further remarks, "individuation is an ontological relationship between entities: what 'individuates' an object, in this sense, is whatever it is that makes it the single object that it is—whatever it is that makes it *one* object, distinct from others, and the very object that it is as opposed to any other thing," (*Ibid*.). Given the latter sense of individuation, Loux points out that for an Aristotelian, the kind to which a concrete substantial entity like the self/person belongs marks it out as a particular that is numerically distinct from other particulars. One way the Aristotelian expresses this matter is by taking universals like *human being*, *dog*, and *oak tree* as individuative universals. The kinds under which concrete particulars like selves/persons fall cut the world up into individual human beings, individual dogs, individual oak trees and so on. In doing that, the kinds provide us with principles for identifying, distinguishing, and counting objects. In light of this, Loux claims that we invoke the kind horse to identify a particular horse, to distinguish different

257

horses and to count horses, saying "one horse, two horses and so on". Moreover, Loux points out that when we do these things we are only recounting the way the kind has partitioned off the world into its instantiations (1998: 130; cf. Moreland 2001: Ch. 7). As Loux sums up:

> The kind constitutes the essence or core being of each of its members; but in virtue of being an instance of its proper kind, a concrete particular can be the subject for attributes—properties—that are external to its core being. So concrete particulars do have a structure that the ontologist can characterize: there is a core being or essence furnished by a kind of and a host of properties that lie at the periphery of that core and, hence, are accidental to concrete particulars. But while they have structure, concrete particulars are not constructions out of more basic things. Since the kind furnishes their essence is an irreducibly unified form of being, concrete particulars are themselves irreducibly unified entities. Their being what they are—human beings…is not to be analyzed in terms of lower-level constituents; they are basic entities (Ibid. 130-131).

So in the Aristotelian picture, it is not up to us to curve reality at its joint. This means that the Aristotelian rejects conventionalism, according to which it is up to people to decide what reality is like (see e.g., Sidelle 1989; cf. Hirsch 1982: 302).

C. *What is the Difference?*

The above view sketched out as the *Neo-Aristotelian View of Substantial Self/Person* is different from Descartes's view of the self in one important respect. As we recall from our discussion in chapter one, for Descartes the self/person has no physical basis, since its essential nature consists in thinking. Descartes also established the existence of the self on an *a priori* basis, for example, through conceiving his own distinction from his body. According to *the Neo-Aristotelian View of Substantial Self/Person*, sketched out above, the self/person has a robust physical basis. By 'robust' I mean the self/person is embodied by *human animal*. In the Aristotelian picture, the self/person has deep unity with the animal body that embodies it. Unlike the Cartesian model that introduces the separation between the self/mind and the body, *the Neo-Aristotelian View of Substantial Self/Person* does not imply any such separation. However, this view maintains that the self/person is distinct from the body or any part of it such as the brain. In this respect, my view is compatible with that of Descartes's, since Descartes also accepts the distinction between the mind and the

body. On the other hand, unlike Descartes who proposed his view of the mind and the body on an *a priori* ground, my view assumes that the history of the self/person (i.e., its origin) as coincident with that of the history of the material body that embodies it. In short, the view defended here assumes that the self/person begins to exist through the natural biological process of conception (i.e., fertilization). Thus, unlike Descartes's, my view roots the self/person in the natural world. According to this view, simply because the self/person has a physical base, it does not mean that the self/person itself is physical.

In fact, my own view is that the self/person is not physical. Being a subject of my mental life, I am in the word of Moreland, a "simple conscious self" (2009: 116). I am also inclined to think that persons survive the destruction of physical death (cf. Gillett 1986: 337-386). In my view, ultimately this is not an issue that philosophy or science can have a last word on it.[142] That is to say that the possibility of surviving the destruction of physical death is a matter that everyone finds out for himself or herself upon death.[143] If as some argue that consciousness may well continue to exist after brain death, then this suggests a live possibility of persons surviving their physical death (see e.g., Kelly, Greyson and Kelly 2007; Beauregard and O' Leary 2007:155). In this case, my view differs from that of Lowe's, since he does not endorse the disembodied existence of persons (see e.g., 2006: 5).

My view is also different from Locke's view of the self. As we discussed in chapter two, despite the problematic nature of Locke's view of the self, he rejects the substantiality of the self. But according to the *Neo-Aristotelian View of Substantial Self/Person*, the self/person is a substantial concrete persisting entity. This view is also different from Hume's view in that for Hume (see chapter three), there is no distinct property bearer such as a substance. Instead the self is a bundle of properties. The *Neo-Aristotelian View of*

_____

[142] See Sorabji's discussion on this issue, 2006: Chaps. 17-19. See also Lewis (1978). Fischer, ed., (1993). Malpas and Solomon, eds., (1995).

[143] In saying this however, I do not mean that, for example, as it is widely believed in Christianity, that there are no good reasons to be certain about one's own survival of physical death. In this thesis, it is not my goal to either enter into such debates or try to settle the issue. For detailed discussion on this issue from the standpoint of Christianity, see e.g., Rea (2009); Davis (1993). That said, however, as I pointed out in chapter one, a theory personal identity must address both the theoretical as well as practical aspects of one's own destiny.

Substantial Self/Person is also different from some of the mainstream views of personal identity. For example, my view is different from the neo-Lockeans view of the self/person, since for the neo-Lockeans, the self/person is not a substantial concrete entity (Parfit 1984: Part III). My view is also different from animalism, since defenders of animalism do not believe in a distinct self/person in addition to the animal physical body (see e.g., van Inwagen 1990). My view is also different from the constitution view of the self/person that relates the notion of the self/person to having a first person perspective (see e.g., Baker 2000). The neo-Aristotelian view defended here, embraces that the self/person has a capacity for first person perspective, but it rejects equating such capacity with personhood as Baker does.

Unlike most of the views surveyed above, the *Neo-Aristotelian View of Substantial Self/Person* grounds a person's identity and individuation in the nature of the entity itself. It takes the commonsense notion of the self/person seriously. Although commonsense ontology needs to be refined in light of the new evidence, it's contribution to our knowledge of the ontology of the self/person is still significant.

D. *Application to Contemporary Issues: Persistence*

According to t*he Neo-Aristotelian View of Substantial Self/Person*, the persistence of persons over time, presupposes that they are substantial entities that endure through qualitative change. Whatever qualitative change(s) a substance undergoes, it is not the case that such changes affect the substance's own persistence over time. However, notice that, here the issue of numerical identity itself turns out to be contentious, if it is seen from the perspective of the composite objects. Composite objects are composed of other objects, which in turn are considered as parts of them (e.g., table, human beings). What can we say about a change such composite objects undergo? For example, the bike my mom had bought me when I was in high school is still in my mom's house in Ethiopia. But over the years, parts of my bike have been replaced. Can we then say that the bike my mom first bought is the very same bike that is now in my mom's house?

Here the natural answer that comes handy is to say that 'yes', the bike that is now in my mom's house is the very same bike that my mom bought. One of the reasons for this is that my bike still retains most of its original parts. But the question remains: how is it that an object (in this case my bike) remains one and the same bike despite having gone through changes of its parts? What if all the parts of my bike had been replaced entirely by new parts? Can we still say that the bike my mom bought me is the very same bike whose parts are now entirely replaced? In an attempt to answer such sorts of questions, philosophical discussions generated puzzles and paradoxes. In this regard, the puzzle of the ship of Theseus and the paradox of the thousand and one cats, among others, are the most familiar (see e.g., Lowe 2002: Ch. 2). But for present purposes, we won't discuss whether such counterexamples pose any serious threat to our main claim that a substance endures as one and the same via qualitative change.

However, a number of non-reductionist philosophers have shown that such counterexamples do not pose a threat to the assumption that a substance endures thorough qualitative change (*Ibid*.). So, when it comes to the non-reductionists' account of personal identity over time, as Baillie rightly remarks, we can answer the persistence question in light of the following considerations: (1) persons are [distinctly] existing basic entities apart from a body and brain as well as sets of mental and physical states; (2) personal identity is a 'further fact,' in terms of being irreducible to these other facts; (3) whether or not identity holds through time is necessarily a determinate matter; and (4) a person's survival is essentially what matters to him or her regarding the future (Baillie 1993a: 8-9).

Notice here that (1) above makes it clear that persons are not identical to (i) their physical bodies; (ii) part of their bodies such as the brain and (iii) the mental and the physical states that they bear. What then are 'persons'? Persons are concrete substantial enduring entities. More importantly, persons or selves are subjects of experience. That is, as we saw in chapter six, persons have a capacity to recognize the fact that they themselves are the subject of their own experience. In this sense, not only do persons possess reflexive self-knowledge, that is, knowledge of one's own identity but they also have knowledge of their

conscious mental states, which is a knowledge of who one is and of what one is thinking and feeling. Moreover, persons are intentional agents who can bring about intentional acts simply in virtue of intending to do such acts.

(2) Above makes it clear that personal identity is primitive in that it does not need any further grounding nor is it explainable in terms of more fundamental terms than itself. In short, personal persistence is an ultimate and unanalyzable fact. If this is true about personal identity, then observable experiences such as memory, spatiotemporal continuity of a person's body or brain can hardly be used to establish personal identity over time.[144] Notice here however that to give a criterion of identity for something is to imply that, that thing is no longer basic. But if personal identity is an unanalyzable fact in a sense that there is nothing more fundamental than itself, then to try to provide a criterion of personal persistence would be to mistakenly assume that personal identity is not basic. But such remarks should not be taken as begging a question against the reductionists, who in fact deny the primitiveness of personal identity (see again chapter five). This is precisely because the very primitiveness of personal identity is the case which can be established independent of our judgment. In short, primitiveness here is constitutive of the very nature of personal persistence. Hence as Lowe nicely explains:

> We can rule out any criterion of identity for a simple substance….For no such criterion would be consistent with the assumed substantial status of the object concerned. One way of demonstrating this then is to point out that if a concrete object is genuinely a simple substance then it could, metaphysically, exist unaccompanied by any other concrete entity apart from certain concrete entities strongly existentially dependent upon itself—and continue to exist, or persist, in this condition. From this it follows that persistence could not depend on the preservation of any relationship involving the other, independently existing concrete particulars, such as other substances….This being so…the diachronic identity of a simple substance cannot be grounded in any equivalence relation defined over objects distinct from itself—which is just to say that it can have no diachronic identity at all (1998: 170).

(4) Above makes it clear that given the non-reductionists' commitment to the numerical identity of personal persistence through time, it is necessarily the case that personal

_____

[144] As we saw, defenders of complex view ground personal identity in one of these things.

262

identity over time cannot be indeterminate. Since numerical identity is an 'all or nothing' relation, there just does not seem to be any room to let personal identity over time to rest on or come in degrees. (4) Above makes it clear that everyone takes his or her future survival seriously. For human persons, the issue of their own personal identity through time is an existentially important matter in a sense that most people take their own future survival seriously. As Noonan (an ardent reductionist himself) puts it, "…unlike other philosophical problems, the nature of personal identity is not merely of interest to professional philosophers, but also a matter of great practical concern to all of us, philosophers and non-philosophers alike. Man has always hoped to survive his bodily death, and it is a central tenet of many religions that such survival is a reality," (Noonan 1989: 1).

So by endorsing (1)-(4) above, the non-reductionists are ontologically committed to taking a human person as an immaterial substance. But for present purpose, the different ways by which the non-reductionists' spell out the notion of an immaterial substance should not concern us here.[145] At the least, the non-reductionists' agree that an immaterial substance is distinct from a purely physical substance such as the body or any part of it such as the brain. So the bearer of all mental states (e.g., belief, desires) is an immaterial substance; whereas the bearer of all physical states (e.g., the firing of neurons) is a physical substance. But the relationship that exists between the immaterial substance and the physical substance is contingent or not absolute. In this case, while the immaterial substance can survive the destruction of the body (e.g., disembodied existence as widely testified in near-death experiences) that embodies it, the same cannot be said of the physical body itself. For example, here we certainly are not implying that the physical body is irrelevant. Rather the point here is that given the foregone discussion, it is the person as an immaterial substance as opposed to the physical body which embodies it (person) that is primitive and ungrounded. On this score, Moreland's remarks hit home:

> I am an all or nothing sort of thing…my identity to myself, in contrast to the degree of similarity between two different objects, does not come in degrees but it, rather, is an absolute fact. My body can gain and lose parts while I retain absolute sameness through such a change…," (Moreland 2009: 113).

_____

[145] For various non-reductionists' characterization of the concept of immaterial substance (see e.g., Lowe 2008; Moreland 2009).

In light of such considerations, therefore, we can provide no necessary and sufficient conditions to establish personal identity over time. Since a person is essentially an immaterial substance, it may well be the case that a person survives the destruction of his/her body while continuing to exist without the body. If this is the correct analysis of personal persistence over time, then a person P_2 at t_2 is the same person as person P_1 at earlier time at t_1, if and only if a person P_2 at t_2 is numerically identical to person P_1 at earlier time at t_1 or at different stages of a person's life. But here by 'stages of a person's life' I do not mean to indicate that a person's identity over time is grounded in temporal parts.[146]

However, here the reductionist might say that the non-reductionists' characterization of the criterion of personal identity is both trivial and uninformative. Instead the reductionist might insist that we should consider either the bodily criterion of identity, the brain criterion of identity or the psychological criterion of identity to establish a non-trivial and informative criterion of personal identity over time. As we already saw in chapter five, there is no agreement among philosophers whether bodily or psychological criterion is the best way to analyse personal identity over time. Thus, it remains very difficult to see how a non-trivial and informative criterion can be established that everyone accepts.

Since we have already made a case for personal identity as primitive and ungrounded, what importance does the evidence question have when it comes to the debate on personal identity? By 'evidence question,' I mean, how we find out whether the same person persists through time. The importance of answering the evidence question has to do with the derivative evidential role it plays for personal identity. For example, in our engagement with other human persons, we usually rely on human organisms as a guide to personal identity. The unique physical characteristics such as fingerprints are a reliable guide to the identity of human organisms, which derivatively can also be used as a reliable guide to personal identity. Thus, in the evidential sense of 'criterion' there is nothing illegitimate in speaking of 'bodily criteria' of personal identity. *Ditto* for the 'memory criterion' of personal identity

---

[146] See for example an excellent critique of this view Lowe (2002, Ch. 3).

(Lowe 1989b: 136-137). Yet from this, it does not follow that the evidential criterion tells us anything metaphysically interesting about what personal persistence consists in. As Olson rightly claims, "If the criminal had fingerprints just like yours, the courts may conclude that he is you. But even if that is conclusive evidence, having your fingerprints is not what it is for a past or future being to be you: it is neither necessary (you could survive without any fingers at all) nor sufficient (someone else could have fingerprints just like yours)," (Olson 2010).

E. *Summary*

In this chapter, I have mainly attempted to sketch out the view I called *The Neo-Aristotelian View of the Self/Person.* While the view needs to be developed further, all I have attempted to do at this stage, was to explain the main aspects of the view. I also applied this view to the issue of persistence. In the next chapter, I will respond to some of the objections I anticipated against this view and will also show how the view could be further enriched through future research.

Chapter Eight

Some Objections and Future Direction

In chapter seven, I sketched out the view I am calling *the Neo-Aristotelian View of Substantial-Self/Person*. I attempted to give a neo-Aristotelian characterization of the nature of the self/person that is rooted in traditional metaphysics with reference to substance ontology. While I fully recognized that the view I sketched out needs to be further developed, I argued that as it stands, it has the potential to contribute to our understanding of the nature of the self/person or, at the very least, in bringing to the table a distinctive way of thinking about the substantiality of the self. In this final chapter, my main aim is to take some objections which I anticipated at several points in the course of my discussion in the last seven chapters. Although it will not be practical to try to respond to every possible objection I might have expected, here I will consider the ones that have immediate relevance and application to the present discussion.

In light of this, this chapter gives me a chance to bring the main thread of argument I have advanced in the last seven chapters, to some kind of closure. In this chapter, I will also point out in what specific way the view I sketched out in chapter seven could be further developed. So the main goal of this chapter is both to conclude in some meaningful way what I have been voicing in defence of a realist account of the ontology of the self/person as well as to point out future directions in which the view could be enriched.

8.1 *Objections*

A. *Bermúdez's Circularity*

As we recall, one of the issues I repeatedly emphasized in this thesis has to do with a non-theory laden conception of our own selfhood/personhood. In chapter one (§ 1.1.1.1), I explained this point in detail using Descartes's *cogito*. I argued, *inter alia*, that Descartes's *cogito* ultimately has to be taken as a *de re* knowledge. Understood this way, *de re* knowledge, which also entails self-consciousness, is rooted in the first-person perspective (cf. O'Brien 2007: Ch.1). One of the central roles of the first person pronoun has to do with

its referentiality (see e.g., Bar-On 2004: Ch. 1). More specifically, when a person comprehendingly uses the first person pronoun, he/she cannot fail to refer to himself/herself. For example, after a very busy day, if I say to my wife, 'I am so tired,' then the 'I' in the statement 'I am so tired' has a clear referent. In this case, the 'I' unmistakably refers to me—the producer of the utterance. Dorit Bar-On calls such utterances—'avowals' which have special security from being prone to referential error (see 2004: Ch.1; also see O'Brien 2007: 4-6; cf. Snowdon in Coliva 2012: 245-256). Although this might seem to be obvious for most of us, some philosophers still want to dispute the referentiality of the first person pronoun (e.g., Anscombe in Cassam 1994: Ch. VIII). I will take up this issue in relation to another objection that I will discuss in due course.

In this section, my focus will be on Bermúdez's circularity objection. Given that my focus so far has been on the first person knowledge of our own selfhood, my claim directly faces Bermudez's circularity objection. As I briefly explained in chapter one § 1.1.1.1, Bermúdez claims that self-consciousness is paradoxical. This paradox arises, according to Bermúdez, when we ponder on what seems to be the common thread that ties the various manifestations of self-consciousness together (1998: 1). Bermúdez claims that this paradox raises the question of how in the first place self-consciousness is possible (*Ibid*.). The paradox consists of two types of circularity, namely explanatory and capacity, respectively (1998:14-24; cf. Sainsbury and Tye 2012: 18). Regarding the explanatory circularity, Bermúdez remarks:

> Any theory that tries to elucidate the capacity to think first-person thoughts through linguistic mastery of the first-person pronoun will be circular, because the explanandum is part of the explanans either directly…or indirectly (1998:16).

Regarding the *capacity circularity* Bermúdez remarks:

> The capacity for reflexive self-reference by means of the first-person pronoun presupposes the capacity to think thoughts with first-person contents, and hence cannot be deployed to explain that capacity. In other words, a degree of self-consciousness is required to master the use of the first-person pronoun (1998:18).

In Bermúdez's account of the paradox of self-consciousness, he also suggests how we can diffuse both forms of circularity. He sums it up as follows:

A nonconceptual content is one that can be ascribed to a thinker even though that thinker does not possess the concepts required to specify that content. Nonconceptual first-person contents are those that fall into this category and that can be specified by means of the first-person or indirect reflexive pronouns. This nonconceptual first-person content offers a way of breaking both forms of circularity (1998:49ff).

Reply: Why should we think that self-consciousness is paradoxical in a way that Bermúdez thinks that it is? Before we answer this question, it is crucial to point out the core assumptions at work behind Bermúdez's proposal of self-consciousness as a paradox. Bermúdez seems to be assuming that the capacity to think first-person thoughts, as well as the capacity for reflexive self-reference, comes through effortful linguistic mastery of the use of the first-person pronoun. In light of this, Bermúdez thinks that one cannot account for the phenomenon of self-consciousness without at the same time running into the problem of circularity, namely the explanatory and the capacity, respectively. Bermúdez suggests that the solution to avoid the circularity problem, in this case, requires us to divorce self-consciousness from the first-person perspective. I am unpersuaded by Bermúdez's remarks for at least three main reasons.

First, Bermúdez is wrong in thinking that self-consciousness can be divorced from the first-person perspective. If such divorce is possible, then how can we elucidate the very notion of self-consciousness? (see e.g., Hamilton 2013: 17). To see the force of this question, we need to understand what the notion of self-consciousness amounts to. Hamilton characterizes it as follows: "self-consciousness is a phenomenon expressed by use of a self-referring device with the properties of the first person," (*Ibid*. 10). At least two things seem to follow from Hamilton's characterization of self-consciousness: (i) the first-person perspective is built into the very notion of self-consciousness. That is to say that one cannot stand without the other. If this is right, then it is hard to see how one can establish a real separation between the two; and (ii) there is no guarantee that one succeeds in giving a non-circular account of self-consciousness (cf. O'Brien 2007: 8-9). If we grant (i) above, then there will always be a residue of an unavoidable circularity at the background of any account of self-consciousness. However, I take such circularity to be benign as opposed to vicious. By 'benign circularity,' I refer to a situation/an explanation that does not lead in a non-

ending/continuous fashion to a series of subsequent explanations. For example, the use of the first person pronoun "I" one uses to refer to oneself is circular, since it presupposes one's prior grasp of what its referent is. So a non-circular account of the use of "I" does not seem to be available. For reasons we shall see shortly, such circularity is benign. Similarly, as Burgess points out, "we need to experience redness to master the concept of redness. We need to experience redness in order to know what redness is," (2008: 223; cf. Alston 1986: 1-30). On other hand, by 'vicious circularity,' I understand a situation/explanation that continuously leads to a non-ending subsequent series of explanations with no real progress. For example, Burgess uses the following example:

1. X is putrid $=_{df}$ x is decayed
2. X is decayed $=_{df}$ x is rotten
3. X is rotten $=_{df}$ x is foul
4. X is foul $=_{df}$ x is putrid.

As Burgess points out, here the series of definitions in (1)-(4) leads us precisely back to our point of departure. Burgess says that the image invoked here is that of a dog chasing its own tail, of the process of definition going nowhere (2008: 216-233). This sort of vicious circularity is not philosophically acceptable for it does not advance a particular argument or thought. By contrast, benign circularity is philosophically acceptable, since it does not stand on the way of a particular argument or thought progression. Hence, self-consciousness is not viciously circular.

So even if for the sake of argument we agree that the two types of circularity that Bermúdez is concerned with should be avoided from our account of self-consciousness, there is another sort of circularity I have in mind which seems to be unavoidable. This is a sort of circularity that is rooted in the phenomenology of self-consciousness itself. For example, the reason why I believe that I am a creature with self-consciousness is because, there is what it is to be like a creature with self-consciousness, which is firmly grounded in my first person awareness. It is such phenomenological experience that primarily enables me to realize that I have the property of self-consciousness. So if I were to give an account of self-consciousness, then my account inevitably would presuppose my prior grasp of my being a creature with self-consciousness. This is an instance of what I shall call: *phenomenological circularity*. As I said earlier, even if for the sake of argument we agree with Bermúdez that

there could be a way to avoid a kind of circularity that he is concerned with, I see no good reason to think that we can do the same with phenomenological circularity. The only scenario we can hope to avoid phenomenological circularity is by abandoning the very notion of self-consciousness, which seems to be rather an extremely unattractive option, to say the least. It is then fair to say that even if Bermúdez succeeds in avoiding his version of the circularity problem, his strategy fails to secure a complete eradication of circularity from our account of self-consciousness. As we shall see, I have doubts about the possibility of avoiding even the types of circularity that Bermúdez has in mind.

Second, if we grant what I earlier called *phenomenological circularity*, then contra Bermúdez, there is nothing wrong in explaining the capacity for reflexive self-reference by means of the first person pronoun. Bermúdez thinks that, since the use of the first person pronoun presupposes the capacity to think thoughts with first-person contents, it cannot be deployed to explain that capacity. My claim is that given my version of phenomenological circularity, we cannot give entirely neutral, that is, a non-circular account of our capacity for reflexive self-reference that does not involve or in some way presuppose the first-person perspective. Hence, either way, we are still stuck with some sort of circularity. If I am right, then Bermúdez's fear of circularity is far-fetched. At this point, it may be asked whether my arguments are intended to show that circular explanations are in principle unproblematic. My answer for such worry would be to say that I do not support explanatory circularity.

However, in the present context, what I am calling phenomenological circularity should not be equated with Bermúdez's explanatory circularity. This is because, from a phenomenological standpoint, when one gives an account of self-consciousness, one does not do that on the basis of any theory. Nor is it that one is trying to give a theoretical account of what self-consciousness is like. Rather in such situations, what one does is simply state the way things are or appear to be to oneself from the standpoint of one's phenomenological experience. In fact, taken in this sense, what Bermúdez calls the 'capacity circularity' can be taken as a species of what I am calling phenomenological circularity. Both cases of circularity are instances of one's prior grasp of self-consciousness.

270

Third, if I am right about the unavoidability of some form of circularity in our account of self-consciousness, then Bermúdez's appeal to a non-conceptual content as a way to diffuse the problem of circularity (whether that be the explanatory or the capacity version) is undermined. Bermúdez thinks that if we specify non-conceptual first-person contents by means of the first person or indirect reflexive pronouns, then we will have a way of breaking both forms of circularity. But this strategy does not seem to work. For example, as I briefly mentioned in chapter one, we can understand non-conceptual first-person content in dispositional terms. Taken this way, the non-conceptual first-person content only needs the right kind of circumstance for its manifestation in terms of a capacity, namely for reflexive self-reference via the use of the first person pronoun. So once we take a dispositional approach to understand non-conceptual first person content, it remains unclear how it can be used as a solution, as Bermúdez suggests, to diffuse the paradox of self-consciousness (assuming that there is such a paradox).

Bermúdez may respond to all of this by saying that my criticisms of his explanatory circularity and capacity circularity miss their target. He might say this because of his conviction that self-consciousness is a paradox. This may well be so. However, if as I tried to show that a non-circular account of self-consciousness is hard to come by, Bermúdez still bears the burden of proof to show us how a complete non-circular account of self-consciousness is possible. I conclude then that the objections I pushed against Bermúdez's circularity claims still stand.

From all this, once again, it seems to follow that our basic sense of our conception of ourselves as selves or persons is primitive—i.e., it is not theory-laden (cf. Bennett and Hacker 2003: 367-377; see also Akeel Bilgram in Coliva 2012: Ch. 12). In this case, I disagree with Harré when he says that self-consciousness is a learned ability to speak and think about oneself (Harré in Peacocke and Gillett 1987: 6; cf. Eilan in Bermudez, Marcel, and Eilan, eds., 1995: 339). I understand self-consciousness, rather to be a dispositional property that a creature of a certain kind has, which, given the right kind of circumstances, a creature of that kind manifests and exercises effortlessly. But this does not mean that disciplines such as neuroscience/cognitive science and psychology do not shed light on the mechanism

of self-consciousness (see e.g., Bermúdez, Marcel, and Eilan, eds., 1-27; Popper and Eccles 1977: 440-442, 476-478; Bennett and Hacker 2003: 12). That is a different matter. My point is rather that the metaphysics of self-consciousness does not depend, at least initially, on the findings/discovery of empirical science.

B. *Tooley-Dennett-Singer's conditions on Personhood/Selfhood*

In this thesis, I defended the substantial conception of the self/person, according to which, *inter alia*, the self is distinct from the capacities it has. I also argued within the framework of Aristotelian substance ontology that a substantial entity such as the self/person does not come in degrees. If we take the Aristotelian approach, it makes little sense to say that the self/person comes in and goes out of existence depending on whether or not it satisfies certain conditions that confer to it the status of selfhood or personhood. However, largely due to Locke's influence, contemporary neo-Lockeans on the one hand and bioethicists on the other hand advanced the ontology of the self/person that stands in sharp conflict with the one defended in this thesis. In this section, my aim is to respond to this objection by drawing insights from the contemporary discussions on the metaphysics of dispositions. My response in this regard is novel, since discussions besetting personhood in the contemporary personal identity literature are not connected, at least in any direct manner, to dispositions.

Despite drawing insights from the contemporary discussion of the metaphysics of dispositions, it is not my goal to enter into the current debates regarding dispositions *simpliciter* (see e.g., Bird 2007; Choi 2006/2012; Mumford in Beebee, Hitchcock and Menzies 2009: Ch. 12). For present purposes, I am only interested in examining the ontology of personhood that is based on certain philosophical assumptions. In this case, I focus on Tooley's, Dennett's and Singer's views of what it means to be a person. Unlike Dennett, both Tooley's and Singer's discussion of personhood takes place in the context of the current debate on the moral status of abortion. In this thesis, it is not my goal to enter into these debates (see e.g., Oderberg 2000).

1. Class Distinction

Central to Tooley's, Singer's and Dennett's notion of personhood is what I shall call the *Lockean person-making* or *person-constituting properties/capacities*. First, in his influential article, *Abortion and Infanticide*, Tooley remarks:

> What properties must something have in order to be a person, i.e. to have a serious right to life? The claim I wish to defend is this: An organism possesses a serious right to life only if it possesses the concept of a self as a continuing subject of experiences and other mental states, and believes that it is itself such a continuing entity….which I will call the self-consciousness requirement (Tooley in Singer 1986: 64; see also Tooley in Steven Luper 2014: Ch.15).

Second, Singer remarks:

> I propose to use 'person' in the sense of a rational and self conscious being, to capture those elements of the popular sense of 'human being' that are not covered by members of the species Homo Sapiens (1993: 87).

Third, Dennett remarks:

> The *first* theme is that persons are *rational beings*….The *second* theme is that persons are beings to which states of consciousness are attributed, or to which psychological or mental or *Intentional predicates,* are ascribed….The *third* theme is that whether something counts as a person depends in some way on an *attitude taken* toward it, a *stance adopted* with respect to it….The *fourth* theme is that the object toward which this personal stance is taken must be capable of *reciprocating* in some way…The *fifth* theme is that persons must be capable of *verbal communication*…The *sixth* theme is that persons are distinguishable from other entities by being conscious in some special way….sometime this is identified as self-consciousness….(in Amélie Rorty 1976: 177-178).

Tooley, Singer and Dennett all agree in characterizing a person as a psychological being with capacities such as rationality, self-consciousness, self-concept and so on. As we saw in chapter two, these are the sorts of capacities that Locke explicitly mentions in his characterization of the notion of a person. Furthermore, Tooley, Singer and Dennett also agree in restricting the class of personhood to those who meet the sorts of conditions listed above. In doing so, they tell us who is (should be) excluded from the class of persons. In this regard, Dennett's list of the excluded group include: (i) infant human beings; (ii) mentally defective human beings; and (iii) human beings declared insane by licensed psychiatrists (Dennett in Amélie Rorty 1976: 175). Singer's list include: (i) non-human animals; (ii) new-born infants; (iii) some intellectually disabled humans (1993: 101). Finally, Tooley's list

include: (i) foetus; and (ii) new-born babies (Tooley in Singer 1986: 60-62). In light of this, they also argue that the terms 'human being' and 'person' should not be used interchangeably, since these terms belong to members that belong to two distinct classes (see e.g., Tooley in Singer 1986: 60-62; Singer 1993: 86-87 and Dennett in Amélie Rorty 1976: 175-176). The question remains: Who is then privileged to be included in the class of persons? Unlike they did with the excluded group, Tooley, Singer and Dennett do not give us any explicit list of the privileged group. Of course, were they to give us such a list, their list would probably suffer from the problem of vagueness, since the Lockean person-making properties come in degrees (cf. Williams 1985: 114). Are we then prepared to say that Tooley, Singer and Dennett succeeded in establishing the class of persons that does not have as its members anyone from the excluded group mentioned above? As I indicated earlier, I will attempt to answer this question within the framework of the metaphysics of dispositions.

2. Objects, Dispositions and their Manifestations

Suppose that a certain object *O* exists. What can we know about *O*? Depending on how it is specified, we may know a good deal about *O*. For example: (i) we may know what dispositions *O* possesses; (ii) we may know under what circumstances those dispositions could be manifested; (iii) we may know what sorts of circumstances may hinder the manifestation of those dispositions and so on. If we grant this, then at least initially, following Martin, we can make the following assumptions. As Martin remarks:

> A particular disposition exists or it does not. You could say of any *unmanifesting* disposition that it straight-out exists, even if it is not, at that time or at any other time, manifesting any manifestation. It is the *unmanifested* manifestation, not the disposition itself, that is the would-be-if or would-have-been-if anything is. There can be a disposition *A* for the manifestation of acquiring a *further* disposition *B* and, of course, disposition *B* need not itself have any manifestation, but disposition *B* can still be unfulfilled terminus of that for which *A* has a specific directedness (1994: 1-2; see also Mumford in Beebee, Hitchcock and Menzies 2009: 269-270).

In the above passage, Martin has made two critical points with respect to the nature of dispositions. First, the absence of the manifestation of certain dispositions does not in any way show that they do not exist. Second, if for whatever reason(s), dispositions are not manifested, then they can be taken as unmanifested manifestations. Consider a china cup.

It has certain dispositions, for example, the disposition to shatter if struck. Here the verb 'struck' stands for what is taken to be a *stimulus condition* and 'shatter' stands for what is understood to be a *manifestation*. So the question remains: will it be the case that every time a stimulus condition is met that we should necessarily expect to see a manifestation of a certain disposition? More specifically, should we necessarily expect to see a china cup shatter when struck? Of course, under normal circumstances, the answer for such questions must be 'yes'. However, consider again a slightly modified scenario whereby a china cup is placed inside a sturdy box, such that when struck, the sturdy box completely absorbs the forceful impact—blocking it from reaching the china cup. In this case, the china cup remains un-shattered. Such is an example that seems to capture Martin's phrase 'unmanifested manifestation.' But is there any other way by which we can take the unmanifested manifestations themselves to be the actual manifestations of a different kind? Following Heil (2012: 120-130), I would say 'yes'. For example, Heil remarks:

> A ball's sphericity endows it with a power to roll. But it is also in virtue of being spherical that the ball has the power to make a concave, circular impression in a cushion, the power to reflect light so as to look spherical, the power to feel spherical to the touch. Talk of single-and-multi-track dispositions or powers is confused from the outset. Powers quite generally are multi-track, if this means that they would manifest themselves differently with different reciprocal partner (2012: 21).[147]

Here Heil is echoing Martin's two points that concern with the nature of dispositions. Martin is an ardent defender of multi-track dispositions. Originally, the term 'multi-track' was coined by Ryle (1963: 114). These are dispositions that are believed to have more than one kind of stimulus condition or manifestation, or both (Bird 2007: 21). Taken this way, powers or dispositions have many reciprocal partners. That means that negative interfering factors such as absences, preventers, antidotes, blockers, inhibitors, etc., will no longer be taken as stopping a certain power from being manifested. This is because such things themselves are dispositions manifesting themselves with various reciprocal partners (Heil 2012: 126-130). Again as Heil remarks:

> What of scurvy and the lack of vitamin C? A living body's healthy condition is a mutual manifestation of myriad finely tuned reciprocal disposition partners. When

---

[147] At this point, following Heil I also want to understand dispositions as powers (2012: 120-130; cf. Mumford in Beebee, Hitchcock and Menzies 2009: 269-270).

one of these is missing, you have a different sort of manifestation, just as you have a different sort of manifestation when you remove one of the cards from a pair of propped-up playing cards. Here, as elsewhere, what you have is not an absence's stepping in and producing a particular kind of effect, but a different collection of reciprocal powers yielding a different kind of manifestation. An absence is not an entity, not something with properties providing it with distinctive powers. But certain kinds of manifestation require appropriately propertied something as reciprocal partners. When these are missing, the result is a different kind of manifestation (2012: 127).

The gist of Heil's point here is that, once we take a multi-track powers model, the manifestation of dispositions is not a one way street, whereby one thing causes another in a linear fashion. On the contrary, the manifestation of powers is the result of causings, i.e., mutual manifestings of various reciprocal partners (*Ibid*. 120). Such considerations help us to have a good grip on Martin's earlier remarks. That is to say that rather than talking about unmanifested manifestation, now we can talk about manifestations *tout court*. The manifestation of powers is multi-faceted in that the apparent absence the manifestation of certain powers does not show that no manifestation is taking place. Rather it only means that a different kind of manifestation is happening. So where does all these leave us? As I already indicated earlier, I will use the hitherto discussion on powers, to respond to *Tooley-Dennett-Singer's conditions on Personhood/Selfhood*.

Reply: If one examines Tooley-Dennett-Singer's conception of personhood by taking a popular functionalist approach, then the conclusion one arrives at on that basis will be radically different, if one were to examine it from the standpoint of ontology. A functionalist approach can be stated as follows: (i) *X* is a person iff *X* manages to play certain agreed upon roles. As I indicated in chapter one § 1.1.2.4, the extension of (i) is open ended allowing artefacts (e.g., robots, corporations, computers) to fall under the sortal term 'person'. Taken from the standpoint of (i), Tooley-Dennett-Singer's conception of personhood may be taken to be unproblematic. However, the underlying assumption behind Tooley-Dennett-Singer's conception of personhood is ontological in nature. The underlying assumption I have in mind can be described as follows: (ii) *X* is a person iff *X* exercises certain properties that are believed to constitute personhood. So my own focus will be on (ii).

Given the multi-track powers model, what is the merit of the class distinction Tooley, Dennett and Singer introduced between 'human being' and 'human person'? Tooley, Dennett and Singer think that foetuses, newly born babies and mentally disabled humans are not human persons. This is because, as they see it, humans with various sorts of mental disability have stopped manifesting powers essential for personhood. On the other hand, foetuses, and newly born babies have not yet begun to manifest powers essential for personhood. In light of such reasoning, Tooley, Dennett and Singer freely assume that the class of 'human beings' is different from that of the class of 'human persons'. However, it remains far from clear how Tooley-Dennett-Singer's class distinction can be plausibly maintained, if examined from the standpoint of the multi-track dispositions' model.

Whether it is in the case of foetus or mentally disabled people, the manifestation of powers is always taking place. In the case of a developing foetus or newly born babies, we can understand personhood in light of the concept of potentiality which is often contrasted with actuality. Although I do not speak French, I have a second order capacity to acquire the first order capacity to speak French. Notice that here both the first and the second order capacities are equal capacities. Simply because I do not speak French now, it does not follow that the second order capacity I have to learn French is not an actual capacity (cf. Frankfurt 1971: 1-2). So the notion of potentiality I suggested above should not be understood as docile, since it is an instance of the manifestation of power (cf. Aristotle *Metaphysics Theta*). In light of such similar considerations, Oderberg remarks, "conception does not bring into existence potential human beings, but an actual human being with a potential to develop, given the right external factors, into a mature human being [human person]", (2000: 21; see also Oderberg 2008: 263-267). Similarly, Puccetti remarks, "a human infant, for example, is not expected to make moral judgments: but since he can enter the human conceptual scheme he is a developing person and is expected to have a moral character of his own some day" (1968: 9; see also Williams 1985: 114). Both Oderberg's and Puccetti's remarks echo the assumption that underlies the multi-track powers.

Moreover, in the case of anencephalic infants (i.e., without brains), we can see powers being manifested albeit in a very different way than we normally expect. *Ditto* with mentally disabled people. For example, if a person becomes totally amnesiac, adopting Tooley-Dennett-Singer's conception of personhood, forces us to conclude that the amnesiac is no longer a person. However, if we adopt the multi-track powers model, the right thing to say would be that, an amnesiac is still a person like the rest of us—despite suffering from a neurodegenerative disease. The only difference between a normal person and that of an amnesiac person lies in the latter, no longer being able to utilize his/her cognitive abilities. This happens due to an entirely different kind of manifestation that led to the loss of the amnesiac person's cognitive abilities. The same is true of comatose patients and other cases of severely disabled people. Despite the current orthodoxy that attempts to divide humans into entirely conventionally based classes, the powers ontology briefly discussed above shows why there is no justification for such a move. In light such considerations, the class distinction suggested by Tooley, Dennett and Singer should be rejected. In rejecting the class distinction, however, one need not thereby also deny the distinction that obtains between a person and the animal body that embodies it. In my own case, as I argued, I maintain such a distinction.

To all of these, Tooley, Dennett and Singer might respond in two ways. First, they may reject the multi-track dispositions model altogether. In that case, it is hard to see what compelling reasons there could be for rejecting it. Moreover, in rejecting the multi-track dispositions model, Tooley, Dennett and Singer owe us a better model. Again it is hard to see what that might be. Second, they might insist that the class distinction between 'human being' and 'human person' is not something they originated. Rather, it is the standard view that is accepted by an overwhelming majority of contemporary philosophers. The standard view invoked here primarily takes a person as a functional concept. But I already argued why the functional concept of a person is deeply unsatisfactory. So the ground for maintaining the class distinction as suggested by Tooley, Dennett and Singer remains to be less than adequate, to say the least.

So what should we say about (ii) above? Recall that (ii) states that: *X* is a person iff *X* exercises certain properties that are believed to constitute personhood. As it stands, (ii) is too restrictive in that it only allows something to be included in the category of persons provided that certain properties are exercised. As we saw, given the powers ontology adopted here, the idea of unexercised powers could be problematic as we saw in the case of Tooley-Dennett and Singer. However, many realists about powers are perfectly happy with the idea of a power existing unexercised. But here is a more pressing ontological issue. How does an object acquire the powers it possesses?

No doubt this question receives different answers in different contexts. But in relation to human persons, we can give a broadly Aristotelian answer for it. In this case, the key assumption that an Aristotelian brings to the table is this. That is, human persons have both natural as well as acquired/learned dispositions. As I understand it, a disposition is natural, just in case an object has it in virtue of being a kind of entity that it is. On the other hand, a disposition is learned just in case an object has it through learning. Two qualifications to keep in mind here, however. First, to have acquired dispositions, sometimes natural powers may be prerequisites. Second, to have acquired dispositions, natural powers may not be prerequisites. For example, we may teach a dog certain skills, say of, riding on the scooter, which is a learned disposition. But no matter how hard we try, we cannot make a dog to be a language speaker, since it lacks a natural disposition. But in the case of *Lockean Person Making Properties,* all humans have the natural dispositions for self-consciousness, self-concept and so on, in virtue of being the kinds of entities that they are. So whether these dispositions are manifested in normal ways or not, no human being is more privileged in having them than the other. If I am right here, then it follows that all humans have equal, i.e., unconditional ontological status. In light of such considerations, (ii) above collapses. I conclude then that such a broadly Aristotelian approach, that grounds the powers objects have in their nature, is far superior to the kind of conventionalist approach adopted by Tooley, Dennett and Singer.

C. Personality Fragmentation and Personhood

Sometimes it has been claimed that some neuropsychological and psychiatric phenomenon poses a serious problem for the continuity of the self (Bermúdez, Marcel, and Eilan, 1995: 6). For example, in his article entitled, *Fragmented Selves: Temporality and Identity in Borderline Personality Disorder*, Fuchs describes the experience of patients with borderline personality disorder. As Fuchs explains, people who suffer from borderline personality disorder lack the capacity to establish a coherent self-concept. They switch from one present to the next, thereby identifying themselves with their momentary state of affect. As a result, such people experience a temporal splitting of the self (2007: 381). Even more so, people who suffer from borderline personality disorder experience:

> A shifting view of oneself, with sharp discontinuities, rapidly changing roles and relationships and an underlying feeling of inner emptiness. There is no sense of continuity over time and across situations, no concept of self-development that could be projected into the future, but only an endless repetition of the same affective states, creating a peculiar atemporal mode of existing. The patients often rapidly change their goals, jobs and friends as well as their convictions and values; they are unable to commit themselves to a set of self-defining values, enduring relationships and long-term aspirations (2007: 382).

What can we say about the implications of such experiences? Do they show that the continuity of the self/person can be called into question? My own answer would be not at all. This is, at least, for three main reasons. First, such stories always presuppose a subject of experience, insofar as they are told directly from the first person perspective of those people who have the experience (cf. Westerhoff 2011: Ch. 3). So when one experiences the loss of a sense of self, that does not by itself imply the discontinuity of the self/person. If that were the case, such stories could not have been told in the first person language. Second, such psychiatric phenomena are indicators of how the self/person can experience complicated cognitive, emotional and psychological pathologies (see e.g., Hobson 1993; Higashida 2007; Grandin 2013; Radden in Gallagher 2011: Ch. 23; Hobson in Gallagher 2011: Ch. 24; Parnas and Sass in Gallagher 2011: Ch. 22; Cavell in Gallagher 2011: Ch. 25). This could be for many reasons. To lesser or greater extent, such experiences are common to people in general. I see, therefore, no compelling reason to think that the fragmentation of the self/person, no matter how worse it could be, that it necessarily implies the

discontinuity of the self. Perhaps, the most convincing reason that supports this conclusion comes from our earlier discussion on powers ontology. As we recall, multi-track powers or dispositions have various reciprocal partners. Given the multi-track powers model, the manifestation of powers takes multiple directions. Even if we have a sense of self, in some situations, we might lose our sense of self. When that happens, the right thing to say is not that I no longer maintain my identity over time. Rather, I should understand my situation in terms of the manifestation of the exact opposite disposition. That is, when I was normal, I used to have a clear sense of myself; but after I experience some sort of cognitive, emotional or psychological crisis, I may experience the loss of a sense of myself. That means again that I am now a subject of a different sort of experience, albeit an undesirable sort.

D. *Unified Consciousness and Personhood/Selfhood*

One of the objections raised against the unity of the self can be linked to the unity of phenomenal consciousness. Here, I will advance my discussion by drawing upon Bayne's article entitled, "The Unity of Consciousness and the Split-Brain Syndrome" (2008: 277-300). In this article, Bayne defends the claim that phenomenal consciousness is unified. That is, subjective experiences such as our visual experiences, auditory experiences, bodily sensations, emotional and mood experiences and the like, other conscious cognitive states, do not occur as phenomenal atoms, rather they occur simultaneously. Put differently, such experiences occur as a conjoint phenomenology, which is to say that there is something it is like to have all of these experiences together, which entails that they are so had. Unified consciousness refers to subjects having total phenomenal state. However, when subjects lack total phenomenal state, such subjects experience disunities in their consciousness.

In this regard, split-brain syndrome can be taken as a paradigm example. The split-brain procedure involves severing the corpus callosum (structure in the brain which connects the left and the cerebral hemispheres) to prevent epileptic seizures from traveling from one hemisphere to the other. The widely held view is that the split-brain patients lack unified consciousness. This is because studies have shown that split-brain patients suffer from behavioral disunities, i.e., inability to give a complete report, when asked to identify objects they are shown. Other studies have shown that such people suffer from representational

disunities, i.e., inability to integrate between the content of the patient's conscious states. For example, as Tye describes:

> A subject, *S*, is told to stare fixedly at the centre of a translucent screen that fills his visual field. Two words are flashed onto the screen by means of a projector located behind, one to the left of the fixation point and one to the right, for example, the words 'pen' and 'knife'. The words are flashed very quickly (for just 1/10 of second) so that eye movements from one word to the other are not possible. This arrangement is one that ensures that the word on the left provides input only to the right hemisphere of the brain and the word on the right provides input only to the left. S is then asked what he saw. *S* shows no awareness, in his verbal responses, of 'pen'. However, if S is asked to retrieve the object corresponding to the word he saw from a group of objects concealed from sight, using his left had alone, he will pick out a pen while rejecting the knives. Alternatively, if *S* is asked to point with his left hand to the object corresponding to the word he saw, he will point to a pen. Moreover, if S is asked to sort through the group of objects using both hands, he will pick out a pen with his left hand and a knife with his right. In this case, the two hands work independently with the left rejecting the knives in the group and the right rejecting the pens (Tye 2003: 109-111).

But, despite such instances of disunities of consciousness, a series of experiments have shown that the split-brain patients retain a fully unified consciousness at all times (see e.g., Bayne 2008: 277-300).[148] Tye also argues that, "split-brain subjects are single persons whose phenomenal consciousness is briefly split into two under certain special experimental conditions, but whose consciousness at other times is unified," (2003: 111-113).

While agreeing with both Bayne and Tye, I would add that the situation we see in split-brain subjects can be understood better in light of the multi-track powers analysis I gave earlier. That is, the incompatible responses split-brain subjects give in an experimental environment can be attributed to the manifestations of different powers or dispositions under different

---

[148] Bayne rightly points out that the notion of the unity of consciousness can be understood in different ways. But following Bayne, here I am only interested in phenomenal consciousness and to see whether split-brain syndrome shows the disunity of phenomenal consciousness. In his book *First Person Plural: Multiple Personality and the Philosophy of Mind*, Braude (1995) gives compelling reasons regarding the presence of psychological unity in people with even the most severe forms of multiple personalities. He rejects the hypothesis that multiple personalities, as well as split brain syndromes, show the presence of more than one self. Such observations further strengthen my own view regarding the continuity of the self over time, despite cognitive and other dysfunctions (see § C above; see also Hacking 1995). For further debates on this issue, see Sperry (1968: 723-733); LeDoux, Willson and Gazzangia (1977: 417-421); Mark Bajakian's multiple person view (2011: 195-204).

circumstances. That said, given that the split-brain patients maintain a complete phenomenal state, it seems totally reasonable to say that even the malfunctioning of the brain (as seen in the split-brain patients) by itself does not seem to necessarily impair one's capacity to have an undivided phenomenal consciousness. In this case, a person with a split-brain syndrome on the one hand and a person without such a disorder on the other hand equally enjoy a unified consciousness. At this point it may be asked: what brain mechanism is responsible in allowing us to have a unified consciousness? Is phenomenal consciousness generated by the brain as whole? If not, is there any specific region in the brain that is responsible for generating phenomenal consciousness? These sorts of questions are not easy to answer. For example, the majority of neuroscientists are well aware of the fact that our cognitive and emotional processes are mediated not by one specific region in the brain but by several brain regions working together in an integrated fashion (Mario Beauregard and Denyse O'Leary, 2007: 47; see also Gallagher, ed., 2011: 4). More precisely, the reason why a single area in the brain does not seem to underlie our phenomenal experiences has to do with what neuroscientists refer to as the *combinatorial capacity problem*. That is, as LaRock remarks, "the possible combinations of features that are represented throughout our lives would seem to far exceed the neural machinery with which we are equipped." (2008: 201-229).

Even if it could well be the case that our experience of unified consciousness may be aided in some way by neuronal subassemblies occupying separate areas in our brain, we are still left with the central question of how the brain, whether taken as a whole or a single part of it, is capable of producing unified consciousness. An eminent neuroscientist, Jeanette Norden states, "...there is no area in the brain where it all comes together. Yet what is my subjective experience? That it's happening to a (sic) "me," and that it is unitary. It is an experience which is happening to me. And we yet have not come up with a better paradigm. We have no idea how the brain is accomplishing this," (Norden, 2007: Part 3, 118). This problem often is referred to as *the binding problem* (*Ibid*. 211). It seems that *the binding problem* presents a serious problem for any view that attempts to undermine the unity of consciousness. One wonders how in the first place unified consciousness is possible. My own view is that the unity of consciousness presupposes its bearer, which I

take it to be the self/person. I will say something briefly about this point later in the chapter. In light of the hitherto considerations, it seems reasonable to say that unified consciousness goes hand in hand with unified self.

E. *The Reference of First Person Pronoun "I" and Personhood/Selfhood*

One of the most popular objections against a realist ontology of the self has to do with the reference of the first person pronoun "I". Ryle once remarked that, "the enigmas that I have in mind all turn on what I shall call the 'systematic elusiveness' of the concept of 'I' (in Cassam 1994: 31). Along similar lines, in his 'On The Phenomeno-Logic of the I,' Castañeda remarks, "many mysteries surround the self, but many of them arise from the fact that a self refers to itself in the first-person way," (in Cassam 1994: 160; see also Evans 1982: 205). Ryle-Castañeda's observations are right in that, 'I' is not a concept to which we can give a ready-made analysis, since it defies the philosophers' usual a 'necessary and sufficient category.' Perhaps, one has to accept the primitiveness of 'I', in a sense that it is a term that cannot be further analysed in terms of anything more fundamental than itself. But most philosophers do not want to settle for the primitiveness of the term 'I'. In the literature, two dominant responses have been given.

The first one is Descartes' answer, i.e., 'I' is a referring expression (see chapter one § 1.1.1.2). Of course, those who follow Descartes's footstep in taking 'I' as a referring expression, do not necessarily embrace his dualism. The stand one takes with respect to whether or not 'I' is a referential term must not necessarily be based on solving the mind-body controversy (cf. Zemach 1972: 70-72). Descartes is not the only one who takes 'I' as a referring expression. For example, Castañeda also takes it as a referring expression. As he remarks, "…a correct use of 'I' cannot fail to refer to the entity to which it purports to refer; moreover, a correct use of 'I' cannot fail to pick up the category of entity to which it refers," (in Cassam 1994: 161). Similarly, Evans remarks, "someone who understands a term as referring to himself must be disposed to regard, as relevant to the truth or falsity of certain utterances involving that term, the occurrence of certain experiences which he is in a position immediately to recognize," (1982: 233; see also Strawson in Cassam 1994: Ch. XII; Kripke 2011: Ch. 10).

On the other hand, there are those philosophers who categorically reject the referentiality of 'I'. For example, Anscombe in her famous article the 'First Person,' argued that, "'I' is neither a name nor another kind of expression whose logical role is to make a reference, *at all* (in Cassam 1994: 154; cf. Wittgenstein 1958: §§ 398, 404, 405 and 410). Similarly, Kenny remarks, "the grammatical error which is the essence of the theory of the self is in a manner obvious when it is pointed out....It will not do, for instance, to say simply that 'I' is the word each of us uses to refer to himself, a pronoun...synonymous with the name of the utterer of the sentence. I is not a referring expression at all, since it is possible to describe one's own action in the third person" (1989: 87; see also Vesey 1973: 24-37).

As we recall, I defended the first person perspective in detail in chapters one and six. For all the reasons given in those chapters, I will not follow the anti-referentialists' stance against the term 'I'. Contra Anscombe, I take the term 'I,' as a logically proper name that refers to a person, who self-consciously uses it—with no chance of reference failure (cf. Strawson in Cassam 1994: Ch. XII; also see Shoemaker in Cassam 1994: Ch. IV; Madell 1981: Ch. 2). The referentiality of 'I' is not rooted primarily in the grammatical structure. As I argued in chapter one and seven, my basic awareness of myself is prior to my coming to grasp with the semantics of the term 'I'. That means that, I need the term 'I,' in order to express what I already know about myself. If this is right, then the referentiality of the term 'I' is rooted in the intention of the user—who uses it to refer to himself or herself. In this case, the user of 'I' is not primarily concerned with the sentence he/she constructs by using the term 'I'. Rather the user is literally thinking of himself/herself as a subject/referent of the term 'I'. Putting aside the question of whether the identity of the referent is material or immaterial self, it remains hard to see what motivates Anscombe's and Kenny's extreme positions when they categorically announce the term 'I' as a non-referring expression. I see no good reason to follow them in this regard.

F. *Neuroscience and Personhood/Selfhood*

In her excellent series of lectures entitled, *Understanding the Brain* (2007: I-III), neuroscientist, Nordan claims that it is not for neuroscience to worry about questions related to the existence of self. Neuroscience does not and cannot tell us whether there is

some kind of quality present in humans that survives the death of both the brain as well as the death of the body. For Nordan, the goal of modern neuroscience is to show and understand how the brain and, ultimately, the mind are related. Notice here that Nordan portrays neuroscience as a neutral discipline on the issue of whether or not the self exists. We may agree with Nordan on this point. It could well be argued that *neuroscience qua neuroscience* simply attempts to understand the structure and the function of the human brain (without holding any position one way or the other on the ontological status of the existence of the self). Yet the immediate worry that pops up here is this: how consistently do neuroscientists avoid the nagging question of human nature which deals with notions such as self, mind, etc.,? The answer seems to be not at all. This is because there are neuroscientists who grapple with the issue of the nature of the self (see e.g., Beauregard 2007).

However, Nordan claims that the central issue in neuroscience is the relation that exists between the brain and the mind. Here we need to know how we are supposed to understand the term 'mind'. We also want to know what role the term 'mind' is supposed to play. In light of this, we can ask if there is something that is true of the mind but not true of the brain and vice versa. More specifically, here our goal is to see if there is a sense in which we can understand 'mind' to be an ontologically distinct feature that refers to thinking persons, as opposed to the material body that embodies such thinking beings. But before we discuss these issues, let us look at Nordan's description of how the brain and the mind are related:

> Perception and cognition—our ability to perceive an external world, to appreciate it; our ability to think about it or have self-reflection; to think, to reason—is also the result of underlying brain processes. The brain gives rise to our subjective sense of experience. We don't know how it does it, but we are sure that it does. When we see, hear, feel, think, or do is the result of underlying brain processes….It's the neural activity of our brains that allows for our ability to have experience in the world. We [can see] from…clinical examples that the corollary of this is true, that is, anything can be taken away with the right brain lesion. You can lose selectively nouns from your speech. You can lose the ability to distinguish dogs from cats or roses from lilies. You can lose the ability to apply morality to ethical decisions. You can lose your sense of self. All of these things can be lost with the right brain lesion (Part 3:165; Cf. Seung 2012).

Nordan clearly distinguishes the various dimensions of our experience. Nordan is right in saying that we experience the external world via perception and sensation. She is also right in pointing out that we have phenomenal consciousness, i.e. we feel something in certain ways (e.g., pain as painfulness). Moreover, we cannot deny the fact that certain neurological cases (e.g. associative *agnosia* in which a person fails to identify shapes) disturb the normal functioning of the brain, resulting in mild to severe disabilities. So far so good! But this is not the end of the matter. Nordan still makes some controversial claims. For example, she assumes that our ability for perception, sensation, reasoning, self-reflection, etc., are all part and parcel of the brain activity. This is rather a very interesting claim with some serious philosophical ramifications. If ultimately every aspect of our experience (as Nordan claims above) arises from the brain activity, then the mind (which is meant to be the bearer of our subjective as well as cognitive experiences) must be identical with the very brain from which it arises. In this case, mind is not a substance in its own right; rather what we call 'mind' is equivalent to higher-order cognitive as well as subjective experiences that humans instantiate. In Nordan's understanding, the firing of neurons in our brain is what is responsible for subjective experiences as well as self-reflective ability that humans have. The bottom line here is that the brain does it all for us (see also, Murphy and Brown 2007; cf. Patricia 1986).

But does it? If we follow Nordan's reasoning here, we must be prepared to endorse that the mind is just the brain and in fact, the converse must also be true. In short, here the identity of the mind and the brain is established. Thus, we do not have to suppose an immaterial entity of any sort, such as the self as the primary bearer of our subjective experiences. Notice that for Nordan, the term 'mind,' is just the brain functioning at a higher-level. But Nordan's identification of the mind as the brain's higher function raises more questions than it answers.

Granting that the brain is a physical substance, which bears physical properties, such as the firing of neurons, metabolic activities, electrical activities, etc., we are still at a loss as to how such activities give us a hint regarding the emergence of the non-physical subjective experience. Here, Nordan's understanding of the role of neuroscience, faces what Levine

287

(1983) calls an '*explanatory gap*' problem. For example, how does the c-fiber firing in one's brain gives rise to the feeling of *painfulness* rather than *itches*? (cf. Kim 2006: 220-224). It is one thing to say that these are the regions in the brain that are responsible for memory or vision or hearing or thoughts or whatever. Yet it is another story altogether to answer the question: why do I seem to experience objects of my perception in distinctive ways? In fact, no amount of the knowledge of the physical brain, whether that be taken at the cellular or molecular level, seems to be capable of giving us access to a person's private subjective experiences. As neuroscientist, Schwartz states:

> Not even the most detailed fMRI gives us more than the physical basis of perception or awareness; it doesn't come close to explaining what it feels like from the inside. It doesn't explain the first person feeling of red. How do we know that it is the same for different people? And why would studying brain mechanisms, even down to the molecular level, ever provide an answer to these questions? (2002: 27).

To my knowledge, Schwartz's question has not been given any satisfactory answer (cf. Lockwood 1989: Chaps. 1-4; see also Beauregard 2007). In light of such considerations, it remains difficult to make sense of Nordan's use of the term 'mind'. Should we understand the term 'mind' as a bearer of the higher-level cognitive/subjective properties or just as a higher-order property *simpliciter*? If understood in the former sense, then the term mind seems to be a non-physical substance in its own right. Alternatively, if the term 'mind' is understood in the later sense, then we are once again back in circles to face the *explanatory gap problem*, i.e., how does the physical brain give rise to the non-physical higher order property? Nordan's construal of the term 'mind' remains ambiguous. Once again such considerations seem to magnify the importance of having a proper ontology of the self/person.

8.2 *Future Direction*

Throughout this thesis, I forcefully argued that the ontology of the self/person needs to be taken seriously. In arguing for this claim, I also explained that the Aristotelian substance ontology plays a central role. In the course of the discussions I advanced in this thesis, there was one particular issue that I would have liked to discuss. The issue I have in mind here has to do with *the emergence of phenomenal consciousness and the self/person*. This is the issue I want to purse in my post-doctoral research. Discussions on the metaphysics of

phenomenal consciousness often tend to be conducted without a direct or explicit link to the ontology of the self/person (see e.g., Chalmers 1996). But if as I argued in this thesis, mental states of whatever variety need a distinct bearer, both for their existence as well as individuation, then the same thing should be true of phenomenal consciousness. That is to say that, phenomenal consciousness too needs a bearer. As we recall, in chapter two, Hume rejected a distinct property bearer. Instead he proposed that perceptions can exist without any bearer. However, as Hume himself later admitted (see e.g., the *Appendix in the Treatise of Human Nature*), he could not establish an independent existence of perceptions without a bearer. Similarly, in talking about phenomenal consciousness, we cannot get rid of the self/person. This is because, the self/person is a subject of phenomenal consciousness. The two are inseparable. That is to say that phenomenal consciousness and the self/person are intimately united. In fact, the former ontologically depends on the latter for its existence (cf. chapter seven § 7.6 A-B). If this is right, then this shows us once again, why we need to take the ontology of the self/person seriously.

A. *The Questions*

If the above remarks are right, then there is good reason to further investigate the nature of phenomenal consciousness. In this case, for example, we can ask whether consciousness is emergent or merely structural property. If it is the former, then we want to know whether it is strongly or weakly emergent. Such considerations lead us to other questions that have to do with the relation between the activity of the brain and consciousness. In light of this, we also want to know whether or not brain activity is the source of phenomenal consciousness. If it is not, then we still want to know whether or not its relation to the brain is based on merely intimate correlations. So to give an adequate account of the nature of consciousness, all aspects of the above questions need to be taken into account. Although the questions are interdependent, they also need to be treated on an individual level (cf. McGinn, 1991 and 1999). More specifically, we can identify the questions in two ways, namely as I shall call them: (1) the *property question*; and (2) the *origin question*. As we recall, (1), attempts to determine the nature of the property of consciousness—i.e., whether it is strongly emergent or weakly emergent. On the other hand, (2) focuses on the origin/source of

consciousness. In most cases (1) is discussed within the context of mental causation, the causal closure principle and the supervenience/exclusion argument (see e.g., Kim 2005: Chs.1 and 2; cf. Gibb, Lowe and Ingthorsson, eds., 2013; Gibb 2010). But the same level of attention has not been given to (2) (see e.g., http://philpapers.org/browse/philosophy-of-mind, Accessed 26 March 2014).

Moreover, (2) is often discussed in relation to the *explanatory gap problem* (see e.g., Levine 1983; 2001). That is: How is it possible for a subjective experience (e.g., pain) to arise from a purely physical/neurobiological system such as the brain? This question is related to what is known as: *the problem of phenomenal consciousness* or the *hard problem of consciousness* (see e.g., Block 2007; Chalmers 1996; Tye, 1995; 2003). Nagel (1974), describes the hard problem of consciousness as: 'what it is like'/what it is like to be in such and such state. But for reasons we shall see, I take both the explanatory as well as the hard problem of consciousness primarily as an epistemological, as opposed to a metaphysical problem. But the key point to note here is that dealing with either problem (i.e., explanatory or hard), does not get to the heart of the problem of (2). This is because, in (2) our primary focus is on figuring out *the origin/source of consciousness*. So it seems that until we settle this issue in (2), making genuine progress in (1) may prove difficult. Since as I pointed out earlier that the self/person and phenomenal consciousness are deeply united, any insight we get into (2) will at the same time shed light on the origin of the self/person. That means that tackling (2) will take the notion of the self/person defended in this thesis one step further. It does this, by giving us insights into the emergence of the self/person (cf. chapter six).

B. *Approach*

In light of this, the best way to approach (2) would be from the standpoint of the diachronic as opposed to the synchronic problem of the origin of consciousness. Here by 'diachronic problem of the origin of consciousness,' I refer to the emergence of consciousness *over time*; whereas in the synchronic case, I mean the emergence of consciousness *at a particular time*. However, whether or not the diachronic approach tackles the origin problem in (2) is a moot point, which needs to be taken seriously. As

far as my own view is concerned, I do not think that it is free from problems. Once we realize these problems, we might need to look for the solution elsewhere.

How might one then go about addressing the questions identified in (1) and (2) above? While many different approaches could be employed in addressing these questions, my own preference is to pursue them within a framework of the view I shall call: *the non-causal spontaneous emergence of phenomenal consciousness.* This view stands in sharp contrast with the currently dominant bottom-up causal account of the origin of consciousness. By contrast, this view adopts a top-down model for the origin of consciousness. I will develop this view by drawing upon insights from the contemporary discussions on the metaphysics of emergence (see e.g., Humphrey, 1997; Lowe, 2008; O'Connor and Yu Wong, 2005; Clayton and Davies, eds., 2006). I will also be linking my view to relevant scientific works. In this case, I will draw upon insights from the origin of life research in biochemistry, where the question of the origin of life takes center stage, which is analogues to the problem of consciousness as stated in (2) above (e.g., Schopf, 1999; Fry, 2000; Davies 1999); quantum mechanics, where the focus is on the causal efficacy of consciousness (e.g., Stapp, 2009; Gao, 2007); and neuroscience, where the emphasis is on neuroplasticity/brain rewiring capacity with implications for top-down mental causation (e.g., Schwartz, Stapp & Beauregard, 2005). The strength of these scientific works lies on attempting to tackle an 'explanatory gap problem' between the physical and the mental. But these scientific works have very little to say about the origin of consciousness *simpliciter*.

In the contemporary emergence debate, *inter alia*, the issue of the causal efficacy of emergent properties takes center stage (see also chapter six). Furthermore, there is a debate concerning whether 'strong emergence,' has any scientific relevance. For example, Bedau (prominent defender of weak emergence), remarks:

> Although strong emergence is logically possible, it is uncomfortably like magic. How does an irreducible…downward causal power arise, since by definition it cannot be due to the aggregation of the micro-level potentialities? Such causal powers would be quite unlike anything within our scientific ken. This not only indicates how they will discomfort reasonable forms of materialism. Their

mysteriousness will only heighten the traditional worry that emergence entails illegitimately getting something from nothing. But the most disappointing aspect of strong emergence is its apparent scientific irrelevance (1997: 377).

Bedau's remarks above raise what I shall call: a *three-pronged objection* against a defender of strong emergence (SE):

(a) The incompatibility of SE with materialism.
(b) The incompatibility of SE with science.
(c) The incompatibility of SE with some causal necessity (i.e., cause-and-effect relations).

If one grants (a)-(c), then obviously one will be forced to dismiss SE. Notice that the underlying assumption of (a)-(c) seems to be that, the emergence of something necessitates some kind of cause. So where such a cause is not available, no genuine property would emerge. Since I take consciousness as an irreducible simple/non-structural emergent property—with its own autonomous causal power unique to that of the base/subvenient level, my view directly faces Bedau's *three-pronged objection*. But as we shall see, Bedau's objection is not insurmountable.

C. *The View*

Having said that, I now discuss the gist of the view I am calling: the *non-causal spontaneous emergence of phenomenal consciousness*. This is the view that attempts to establish the origin of consciousness as a strongly emergent property. Here following Moreland, I understand finite consciousness to be an irreducible emergent property—i.e., a simple non-structural property that is unique to the "emergent" level of reality. Taken this way, an emergent property is not captureable, for example, in terms of the spatio-temporal, physical, causal relations among the states as well as parts at the lower level of reality. This means that, an emergent property being a *sui generis*, i.e., a new/novel kind of property, is different from and not composed of the parts, properties, relations, and events at the subvenient level (Moreland 2008: Chs.1-2). One way to make sense of this view is by distinguishing it from a structural property—which is constituted by the parts, properties, relations, and events at the subvenient level—that is identical to a configuration pattern among the subvenient entities. Unlike an

emergent property, a structural property is not a new kind of property (*Ibid*.: 15). Understood this way, a structural property (e.g., solidity) can be both reducible to and identifiable with a purely physical property (see e.g., Jackson 1998: ch.1). The key aspect of the above characterization of phenomenal consciousness as an emergent property has to do with its emphasis on its non-causal origin. That is, according to the view I am proposing, there is no *causal-link* between a strongly emergent property (in this case, consciousness) and its base level—although the base level (as far as we know) serves as a sufficient condition (but not as a necessary condition) for its emergence (cf. Levine 2001: 7).

In this regard, my view differs from orthodox accounts of the emergence of phenomenal consciousness—which requires some sort of causal link to its base level, at least, initially (see e.g., Chalmers in Clayton and Davies, eds., 2006: 244; O'connor 2000: 110-121; Sperry 1969; cf. Searle 1992: Ch.5; Murphy and Brown 2007). Those who think of the emergence of consciousness in causal terms, often have in mind, brain complexity as a source of such causality. However, the causal proposal faces two serious empirically supported objections. I shall call these objections: the *gametes problem* and the *complexity problem,* respectively. The first objection deals with conception (i.e., fertilization) and the impossibility of tracking down/spotting the origin/emergence of consciousness at any point during the process of the development of a foetus. This objection effectively poses the question: 'when does consciousness begin?'—which is very much similar to another question both bioethicists and medical experts struggle with. That is, 'when does life begin?' The second objection focuses on demonstrating that, no amount of analysis of the physical basis of complex brains will unfold anything substantial about the origin of consciousness (cf. McGinn 1999). Both of the two objections above target the diachronic (i.e., over time) account of the origin of consciousness mentioned earlier.

As we recall, the origin question in (2) lies entirely at the heart of the account of the non-causal emergence of consciousness. If this is right, then the emergence of consciousness has no informative analysis. Put differently, it is a fundamental/primitive

fact in the natural world—which cannot be further analysable in terms of more fundamental facts than itself. That means that the emergence of consciousness is an ontological add-on, i.e., has no ordinary causal explanation for its emergence (cf. Moreland 2003:208). However, considering consciousness as a primitive fact faces objection. For example, philosophers like Bedau claim that accepting phenomenal consciousness as primitive is 'too much like magic.' For now, two things can be said in response. First, if phenomenal consciousness is like magic, and by 'magic,' if it is meant something hard to make sense of its origin, then that only speaks about our own epistemic limitations. In that case, I cannot see the force of the objection. By contrast, if by 'magic' if it is meant to show a situation where someone simply postulates how things are due to lack of any other way to account for it, then the primitiveness of phenomenal consciousness can be called into question. But my own proposal of the primitiveness of phenomenal consciousness will not be prone to such an objection, since I do not take the origin of consciousness as a brute postulation. Rather, I take it to be an objective feature of reality.

Second, the 'too much like magic' objection seems to assume that everything has to have some form of empirically confirmed analysis or explanation. At the least, the assumption that is buried beneath the objection is that, if something is not explainable on the basis of some sort of generally agreed upon criterion of explanation, then any other explanation that is not constrained by such parameters has to be deemed as mysterious. I remain unpersuaded by such an objection. For one thing, we believe in so many things despite the fact that we lack any knowledge regarding their origin. Here is one example. That is, no one knows when or how the first living cell emerges; but that does not prevent us to talk about or understand its nature (see e.g., Rana 2011: 102-112; Rana and Ross 2004). No one has a clue as to how to define life itself. But that does not mean that we cannot characterize it (*Ibid*. 24-31). If there are things that reflect an objective feature of reality, I see no good reason not to accept them at face value and live with them. In this case, they only sound more 'like magic' for finite creatures like us. However, I do not claim that this is a decisive answer against the objection considered here. But at the least, I tried to unlock the motivation behind it. I find the motivation deeply unsatisfying. The objection considered here is very much similar to another objection raised against a realist conception of the

ontology of the self. But in chapters one, six and seven we saw why such objections against the primitiveness of the concept of the self/person do not hold water. It seems like reality contains many such primitives and our job is to try to distinguish those that really are primitive. In saying this, however, I do not mean to suggest any criterion we should use. That is not my present goal.

Having discussed the non-causal approach to investigating the nature of consciousness, we also need to think about how once it emerges, consciousness enters into what I shall call: the domain of causal interaction—with brain states. It is at this point that we can plausibly grant both the top-down causal influence of phenomenal consciousness on the base level on the one hand, and the causal influence of the base level (i.e., bottom up) on consciousness as well as other mental states on the other. It is also at this point that we can properly address both 'the explanatory gap problem' and 'the hard problem of consciousness'—each of which I briefly mentioned earlier. In short, it is at this point, where we face the age-old problem of the link between mental properties and physical properties.

Finally, we also need to explain in what manner phenomenal consciousness continues to co-exist with its physical base (e.g., brain). To show this, we must use the notion of correlation and its relation to causation. The key point here is that correlation is not causation (see for details: Maudlin 1994/2002: Ch.5). As neuroscientist Jones nicely sums up:

> The simple act of finding neural correlates for certain behaviours or attitudes provides few, if any, insights into causative factors. Even if a certain brain structure were strongly associated with…[some sort of subjective] experience, this says nothing about whether the structure generates that experience. Simply because brain region "R" is active when behaviour "B" is undertaken does not mean that changes in "R" cause "B" to take place. The opposite, in fact, could be the case, in that when an individual displays behaviour "B," brain region "R" is modified, and if this occurs sufficiently often, there are significant changes to "R." Yet again, the interplay between "R" and "B" may be so close that the only tenable conclusion is that there is no definitive causative factor-the one feeds on the other, (Jones, 2010: 125; also see Schwartz and Begley 2002; Robinson 2007).

In light of Jones's remarks above, we can see that from the temporal/physical standpoint, the correlation between phenomenal consciousness and brain state is necessary. For example, a normal functioning of brain is necessary for normal function of mental state. But from this nothing follows to the effect that such correlations are metaphysically necessary. It may very well be the case that consciousness can continue to exist in the absence of brain or brain function (see e.g., Edward F. Kelly et al., 2007). Such considerations show the utter implausibility of the reductive materialist views that deem consciousness as nothing but brain function (e.g., Place, 1956; Prinz, 2012; Dennett, 1991; Crick and Koch, 1990; Crick, 1994).

D. *Summary*

I introduced a three-stage analysis of the view of consciousness I proposed. The first one has to do with a *non-causal stage.* The second one has to do with a *causal stage.* Finally, the third one has to do with a *correlation stage*. Putting things in this way clears up a lot of confusion that is prevalent in the contemporary debate on the nature of consciousness. Often time philosophers and neuroscientists claim that they are making inroads on the origin problem. But as it turns out, they are only confusing their achievement in phase two (which is a causal phase) with that of phase one (i.e., a *non-causal phase*). The origin question in (2) as pointed out earlier still remains unresolved. There is also wide spread confusion with respect to the third phase in that many theorists mistake correlation for causation. So my three-phase analysis of consciousness clearly identifies the three features that make up the nature of consciousness. That said, in light of what has been said so far, Bedau's *three-pronged objections* can now be laid to rest or at least the foregone discussion has shown us how we can effectively meet such objections.

8.3 *Conclusion*

I began this thesis with what I called the *substantial-self/person question*. In chapter one, I gave an analysis of this question within the framework of the Aristotelian substance ontology. Similarly, adopting Descartes's notion of the entityhoodness of the self and Boethius's notion of a person, I linked the two notions together with the notion of a

substance. In doing so, in the first half of chapter one, I established the conditional claim that I defended in this thesis. That is: *If I am a self/person then I am a certain kind of substantial entity*. I then located this claim in the second half of chapter one, within the context of the contemporary personal identity debate. In chapters two and three, I discussed how two important historical figures, Locke and Hume, respectively dealt with *the substantial-self/person question*, *substance* and *personal identity*. In chapter four and five, I tried to assess the contemporary personal identity debate against the historical backdrop of Locke's and Hume's views of substance and personal identity.

My central claim in this thesis has been that to make a real progress in our understanding of the ontology of the self/person, we need to take metaphysics seriously. In light of this, I expressed my skepticism regarding the effectiveness of the contemporary personal identity discussions. My skepticisms in this regard stem from two main observations. The first observation has to do with the contemporary personal identity theorists' excessive emphasis on the issue of the criterion for personal identity. The second observation has to do with the issue of the methodology of thought experiments in the contemporary personal identity discussions. In light of both of these observations, I argued that the proper subject matter which I called the *Aristotelian Question* (i.e., 'what is X?') has been side stepped. Throughout this thesis, therefore, I insisted that this trend has to be reversed. That is, I argued that the *Aristotelian Question* must be restored to its proper place. By this I meant that our focus on personal identity discussion must be, first and foremost, to grasp the nature of the kind of an entity we call a self/person. I also argued that without taking ontology seriously this question cannot be adequately pursued. In this case, I strongly suggested that traditional metaphysics, as conceived by Aristotle, provides the most preferable framework to pursue the question of the ontology of a self/person.

In light of this, in chapter six, I argued against the naturalistic ontology that tries to undermine the traditional metaphysics. Alternatively, I attempted to show how we can have a robust understanding of the ontology of a self/person if we take metaphysics seriously. I defended this claim further in chapter seven, where I sketched out the view I called the *Neo-Aristotelian View of Substantial Self/Person*. In chapter seven, I argued that one of the

advantages of traditional metaphysics is that it respects our common-sense conception of ourselves. At the background of everything I said in this thesis in defence of a realist conception of a self/person, the common-sense conception is assumed. I argued that common-sense conception of ourselves is the proper starting point in thinking about the ontology of a self/person. I argued that theory-laden understanding of the ontology of a self/person puts the cart before the horse. In this chapter, I tried to show that our common-sense conception of ourselves gives us an unadulterated initial insight into how we naturally view ourselves. For example, most people view themselves as persisting entities over time. Most people also have strong convictions about the possibility of surviving in some form their own bodily destruction. Most people naturally view themselves as agents who can engage in intentional actions. I assumed that such pre-philosophical convictions people have about themselves must be taken seriously as opposed to being explained away. Taking such things into consideration, in this chapter, I sketched out *the Neo-Aristotelian View of Substantial Self/Person*. I located this view entirely within the framework of Aristotelian metaphysics with reference to substance ontology. Although the view I sketched out needs to be developed further, I tried to show how, at the least, Aristotelian substance ontology can be put to a good use. Given that Aristotle's theory of substance has not been given the attention and the place it deserves in the contemporary personal identity mainstream literature, I consider my attempt to be a step in the right direction.

In this final chapter, I picked out a handful of objections often raised in the literature to challenge a realist conception of an enduring/persisting self/person. I tried to show how these objections do not pose any serious problem for the view of the self/person defended in this thesis. In this chapter, I tried to show how the metaphysics of dispositions or powers ontology provides us with an excellent framework to respond to most of the objections I considered. Finally, I ended this chapter by mapping out a specific area of future research, the emergence of phenomenal consciousness and the self/person. This will be a second phase of my research. I hope in this thesis, I made a strong case for why we need to take traditional metaphysics as well as substance ontology seriously in order to come up with a robust ontology of the self/person.

Bibliography

Abelson, R. (1977). *Persons: A study in philosophical psychology.* London, England: The Macmillan Press. Ltd.

Ackrill, J.L. (Trans.). (1963). *Aristotle's categories and de interpretatione.* Oxford: Clarendon Press.

Ackrill, J.L. (Ed.) (1987). *A new Aristotle reader.* Princeton, NJ: Princeton University Press.

Alston, W.P. (1977). 'Self-intervention and the structure of motivation.' In Theodore Mischel (Ed.), *The self: Psychological and philosophical issues.* Rowman & Littlefield.

Alston, W. P., and Bennett, J. (1988). 'Locke on people and substances'. *The Philosophical Review*, 25-46.

Alston, W. P. (1986). 'Epistemic Circularity.' *Philosophy and Phenomenological Research*, 47.

Anscombe, G.E.M. (1994): 'The first person.' In Cassam, Q. (Ed.). *Self-knowledge.* Oxford: Oxford University Press, pp. 140 – 159.

Antonietti, A., Corradini, A. and Lowe, E.J. (Eds.) (2008). *Psycho-physical dualism today: An interdisciplinary approach.* Lanham, MD: Rowman & Littlefield Publishers, Inc.

Anstey, P. (2000). *The philosophy of Robert Boyle*. London: Routledge.

Armstrong, D.M. (1980). *The nature of mind.* Sussex, UK: The Harvester Press Ltd.

Armstrong, D.M. (1989). *Universals: An opinionated introduction.* Boulder, CO: Westview Press.

Armstrong, D. M. (1999). *The mind-body problem: An opinionated introduction*. Boulder, CO: Westview Press.

Atkins, K. (Ed.). (2005*). Self and subjectivity.* Oxford, UK: Blackwell Publishing.

Atherton, M. (1998). 'Locke and the issue over innateness.' In V. Chappell, (Ed.), *Locke.* Oxford: Oxford University Press.

Audi, R. (1998). *Epistemology: A contemporary introduction to the theory of knowledge.* London, England: Routledge.

Audi, R. (1998). 'Moderate intuitionism and the epistemology of moral judgment.' *Ethical Theory and Moral Practice*, *1*(1), 15-44.

Augustine, *De civitate dei, XI.*

Augustine, *De civitate dei, XIII.*

Augustine, *De trinitate, Book X.*

Aune, B. (1985). *Metaphysics: The elements.* Minneapolis, MN: University of Minnesota Press.

Ayer, A. J. (1936). *Language, truth, and logic*. London, Victor Gollancz.

Ayer, A. J. (1954). *Philosophical essays.* London, England: Macmillan.

Ayer, A.J. (1963). *The concept of a person and other essays.* London, England: Macmillan & Co. Ltd.

Ayer, J. L. (1967). *Metaphysics and common sense, essays on knowledge, man as a subject for science, chance, philosophy and politics, existentialism, metaphysics, and a reply to Austin on sense-data theory.* London, England: Macmillan.

Ayer, A.J. (1980). *Hume.* Oxford, UK: Oxford University Press.

Ayers, M. (1991). *Locke: Epistemology and ontology.* London, UK: Routledge.

Ayers, M. (1990). *Locke*, vol. 2. London: Routledge.

Baillie, J. (1993a). *Problems in personal identity.* New York, NY. Paragon House.

Baillie, J. (1993b). 'What matters in survival?' *The Southern journal of philosophy*, *31*(3), 255-261.

Baillie, J. (2000). *Hume on morality.* London, England: Routledge.

Baillie, J. (2003). *Contemporary analytic philosophy.* Prentice Hall.

Baker, L.R. (2000). *Persons and bodies: A constitution view.* Cambridge, UK: Cambridge University Press.

Bajakian, M. (2011). 'How to count people.' *Philosophical Study,* 154:185–204.

Barger, B. (1976). *Locke on substance. And Boyle, R. Origin of forms and qualities (The theoretical part)*. Manhattan Beach, CA: Sheffield Press.

Bar-On, D. (2004). *Speaking my mind: Expression and self-knowledge.* New York, NY: Oxford University Press.

Baumgartner, S., Heisenberg, T. and Krebs, S. (2013). *Metaphysics or modernity? Contributions to the Bamberg Summer School 2012. Bamberg, Germany: University of Bamberg Press.

Bayne, T. (2008). 'The unity of consciousness and the split-brain syndrome.' *Journal of Philosophy,* 105 (6), pp. 277-300.

Beauchamp, T.L. (Ed.). (2000). *David Hume: An enquiry concerning human understanding.* Oxford, UK: Clarendon Press.

Beauchamp, T.L. (Ed.). (1998). David Hume: *An Enquiry concerning the Principles of Morals*. Oxford: Oxford University Press.

Beauregard, M., O'Leary, D., and Lawlor, P. (2007). *The spiritual brain.* San Francisco, CA: HarperCollins Publishers.

Bennett, J. (1971). *Locke, Berkeley, Hume: Central themes.* Oxford, UK: Clarendon Press.

Beebee, H., C. Hitchcock & P. Menzies. (Eds.) (2009). *The Oxford Handbook of Causation.* Oxford: Oxford University Press.

Bennett, J. and Alston, W. (1988). 'Locke on people and substance'. *Philosophical Review*, 97, pp. 25-46.

Benedict, S. (2014). Unpublished paper. Durham University.

Bennett, M.R., Hacker, P.M.S., and Bennett, M.R. (2003). *Philosophical foundations of neuroscience.* Oxford: Blackwell, pp. 355-378.

Berkeley, G. (1713*). Three dialogues between Hylas and Philonous.* Edited by Adams, R.M. Cambridge, UK: Hackett Publishing Company.

Bermúdez, J.L. (1998). *The paradox of self-consciousness.* London, England: The MIT Press.

Bermúdez, J.L. (2002). 'The sources of self-consciousness.' *Proceedings of the Aristotelian Society,* New Series, Vol. 102 pp. 87-107.

Bermúdez, J.L., Marcel, A. and Eilan, N. (Eds.). (1995). *The body and the self.* Cambridge, MA: The MIT Press.

Bird, A. (2007). *Nature's metaphysics: Laws and properties.* Oxford, UK: Oxford University Press.

Brie G. and Lawrence A. S. (Eds.) (2007). *Arguing about the mind*. London: Routledge.

Braude, S. E. (1995). *First-person plural: Multiple personality* and *the philosophy of mind.* Rowman & Littlefield.

Black, M. (1952). 'The identity of indiscernibles.' *Mind*, *61*(242), 153-164.

Blackburn, S. (1994, 1996, 2005). *The Oxford dictionary of philosophy. (2ⁿᵈ Edition).* Oxford, UK: Oxford University Press.

Blackburn, S. (1997). In J. Dancy (ed.), *Reading Parfit.* Blackwell, pp. 180-201.

Blackburn, S. (1999). *Think*. Oxford: Oxford University Press.

Block, N. (2007). *Consciousness, function, and representation: Collected papers*. Cambridge: MIT Press.

Bonevac, D. (2003). *Deduction: Introductory symbolic logic. (2ⁿᵈ Edition).* Oxford, UK: Blackwell Publishing.

Bolton, R. (1994). *Person, Self, and Identity*. London: Minerva Press.

Boethius (1981). *The Theological Tractates and the Consolation of Philosophy*. Edited by H. F. STEWART and E. K. RAND. London: W. Heinemann.

Boethuis. *Categorias Aristotelis libir quattuor*. Migne PL 64.

Bradley, F.H. (1908). *Appearance and reality: A metaphysical essay.* New York, NY: The Macmillan Company.

Bricke, J. (1980). *Hume's philosophy of mind.* Princeton, NJ: Princeton University Press.

Brody, B. A. (1980). *Identity and essence*. Princeton: Princeton University Press.

Brook, A. (2012). 'Review of 'The unity of consciousness' by Tim Bayne', *Australasian Journal of Philosophy,* 90:3, pp. 599-602.

Brown, J. (2011). 'Thought experiments, intuitions and philosophical evidence.' *Dialectica*, *65*(4), 493-516.

Brown, J. R. and Fehige, Y. *Thought Experiments*. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. 2008.

Burgess, J.A. (2008). 'When is circularity in definitions benign?' *The Philosophical Quarterly* Vol. 58, No. 231, pp. 214-233.

Campbell, K. (1990). *Abstract particulars.* B. Blackwell.

Carter, W. (1990). *The elements of metaphysics*. Temple University Press.

Capaldi, N. (1975). *David Hume: The Newtonian philosopher.* Boston, MA: G.K. Hall & Co.

Care, N.S. and Grimm, R.H. (1969). *Perception and personal identity: Proceedings of the 1967 Oberlin Colloquim in Philosophy.* Cleveland, OH: The Press of Case Western Reserve University.

Carnap, R. (1950). 'Empiricism, semantics and ontology.' *Revue Internationale de Philosophie, 4,* pp. 20-40. Reprinted in the supplement to *Meaning and necessity: A study in semantics and modal logic,* enlarged edition (University of Chicago Press, 1956).

Carr, B. (1987). *Metaphysics: An introduction.* London, England: Macmillan Education Ltd.

Carr, D. (1999). *The paradox of subjectivity: The Self in the Transcendental Tradition.* Oxford, UK: Oxford University Press.

Carruthers, P. (1986). *Introducing persons: Theories and arguments in the philosophy of mind*. SUNY Press.

Carruthers, P. (2004). *The nature of the mind: An introduction.* New York, NY: Routledge.

Cassam, Q. (1994). (Ed.). *Self-knowledge.* Oxford, UK: Oxford University Press.

Cassam, Q. (1997). *Self and world.* Oxford, UK: Clarendon Press.

Castañeda, H.N. (1974). *The structure of morality.* Springfield, IL: Charles C. Thomas,

Cavell, M. (2011). 'The self: Growth, integrity, and coming apart.' In Shaun Gallagher (ed.), *The Oxford Handbook of the Self.* Oxford, UK: Oxford University Press.

Chalmers, D.J. (1996). *The conscious mind: In search of a fundamental theory.* Oxford, UK: Oxford University Press.

Chalmers, D.J. (2012). *Constructing the world.* Oxford, UK: Oxford University Press.

Chalmers, D.J. (1995). 'Facing up to the problem of consciousness', *Journal of Consciousness Studies* 2:3, pp.200-219.

Chalmers, D., Manley, D., and Wasserman, R. (Eds.) (2009). *Metametaphysics*, Oxford: Oxford University Press.

Chappell, V. (Ed.). (1998). *Locke.* Oxford: Oxford University Press.

Chappell, V. (1989). Locke and relative identity. History of Philosophy Quarterly, 6 (1), pp. 69-83.

Chisholm, R.M. (1969). 'On the observability of the self', *Philosophy and Phenomenological Research,* Vol. 30, No. 1, pp. 7-21.

Chisholm, R.M. (1976). *Person and object: A Metaphysical Study.* London, England: George Allen & Unwin Ltd.

Chisholm, R.M. (1989). *On metaphysics.* Minneapolis, MN: University of Minnesota Press.

Chisholm, R.M. (1996). *A realistic theory of categories: An essay on ontology.* Cambridge, UK: Cambridge University Press.

Churchland, P.M. (1988). *Matter and consciousness: A contemporary introduction to the philosophy of mind. (Revised Edition).* London, England: The MIT Press.

Clyton, P. and Davies S.D (Eds.). (2006). *The Re-Emergence of Emergence*. Oxford: Oxford University Press.

Code, A. (1986). 'Aristotle: Essence and Accident.' In R.E. Grandy & R. Warner (Eds.), *Philosophical Grounds of Rationality*. Oxford: Clarendon Press.

Code, A. (1978). 'What is it to be an individual.' Journal of Philosophy, 75 (11), pp. 647-648.

Cohen, S.M. (2009). 'Substances.' In G. Anagnostopoulos (ed.), *A Companion to Aristotle*. Oxford, UK: Blackwell, 197-212.

Cohen, S.M. (2014). [Online] 'Aristotle's metaphysics,' *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.), Available from: http://plato.stanford.edu/archives/sum2014/entries/aristotle-metaphysics/.

Coliva, A. (Ed.) (2012). *The self and self-knowledge.* Oxford, UK: Oxford University Press.

Coleman, S. (2000). 'Thought experiments and personal identity.' *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition,* Vol. 98, No. 1, pp. 53-69.

Connell, R. J. (1988). *Substance and modern science.* Center for Thomistic Studies, University of St. Thomas.

Cook, J.W. (1968). 'Hume's scepticism with regard to the senses.' *American Philosophical Quarterly,* Vol. 5, No. 1, pp. 1-17.

Corrandini, A., Galvin, S. and Lowe, E.J. (2006). *Analytic philosophy without naturalism.* London, UK: Routledge.

Corradini, A. and O'Connor, T. (2010). *Emergence in science and philosophy.* New York, NY: Routledge.

Cottingham, J. (1993). *A Descartes dictionary.* Oxford, UK: Blackwell Publishers.

Cottingham, J., Stoothoff, R. and Murdoch, D. (1984). *The philosophical writings of Descartes.* (Vol. II). Cambridge, UK: Cambridge University Press.

Cottingham, J., Stoothoff, R. and Murdoch, D. (1985). *The philosophical writings of Descartes.* (Vol. I). Cambridge, UK: Cambridge University Press.

Copan and Moser (Eds.). (2003). *The rationality of theism*. London: Routledge.

Coventry, A., and Kriegel, U. (2008). 'Locke on consciousness.' *History of Philosophy Quarterly*, 221-242.

Craig, W.L., and Moreland, J.P. (Eds.). (2002). *Naturalism: A critical analysis.* London, England: Routledge.

Crick, F. (1994). *The astonishing hypothesis: The scientific search for the soul*. New York: Scribners.

Crick, F. and C Koch, (1990). 'Toward a neurobiological theory of consciousness', *Seminars in the Neurosciences*, 2, pp. 263-275.

Daly, C. (2013). *Philosophy of language: An introduction.* London, England: Bloomsbury.

Dancy, J. (1987). *Berkeley: An introduction.* Oxford, UK: Basil Blackwell.

Dancy, J. (Ed.). (1997). *Reading Parfit.* Oxford, UK: Blackwell Publishers.

De Clercq, R. (2005). 'A criterion of diachronic identity based on Locke's principle.' *Metaphysica*, 6 (1), pp. 23-38.

DeGrazia, D. (2006). 'On the question of personhood beyond Homo sapiens.' In Singer, P. (Ed.). *In defense of animals. The second wave*, 40-53.

Dennett, D. (1976). 'The identities of persons.' In Amelie Oksenberg Rorty (Ed.), *The identities of persons.* University of California Press.

Dennett, D.C. (1984). *Elbow room: The varieties of free will worth wanting.* MIT Press.

Dennett, D.C. (1991). *Consciousness explained.* Boston, MA: Little, Brown and Company.

Dennett, D. C. (2003). *Freedom Evolves*. New York: Penguin Books.

Doepke, F.C. (1996). *The kinds of things: A theory of personal identity based on transcendental argument.* Chicago, IL: Open Court.

Ducasse, C.J. (1951). *Nature, mind, and death.* La Salle, IL: The Open Court Publishing Company.

Dummett, M. (1973/1981). *Frege: Philosophy of Language*. Second Edition. Cambridge, MA: Harvard University Press.

Dummett, M. A. (1981). *The interpretation of Frege's philosophy.* London, UK: Gerald Duckworth & Company Limited.

Durrant, M. (1993). *Aristotle's de anima in Focus*. London: Routledge.

Elder, C.L. (2004). *Real natures and familiar objects.* Cambridge, MA: The MIT Press.

Elster, J. (Ed) (1985). 'Rationality, morality, and collective action.' *Ethics*, 136-155.

Evans, G. (1978). *Can there be vague objects?* Oxford, UK: University College.

Evans, G. (1982). *The varieties of reference.* Oxford, UK: Oxford University Press.

Frankfurt, H. G. 'Freedom of the will and the concept of a person.' *The Journal of Philosophy*, vol. 68, No. 1, pp. 5-20.

Feigl, H. and Sellars, W. (Eds.) (1949). *Readings in philosophical analysis.* New York: Appleton-Century-Crofts, Inc.

Feldman, R. (2003). *Epistemology*. Prentice Hall.

Fine, K. (1995). 'Ontological dependence.' *Proceedings of the Aristotelian Society,* New Series, Vol. 95, pp. 269-190.

Fischer, J. M. (1993). *The metaphysics of death*. Stan., California: Stanford University Press.
Flage, D.E. (1990). *David Hume's theory of mind.* London, England: Routledge.

Flew, A. (1951). 'Locke and the problem of personal identity.' *Philosophy*, *26*(96), pp. 53-68.

Flew, A. (1961). *Hume's philosophy of belief: A study of his first Inquiry.* London, England: Routledge.

Flew, A. (1986). *David Hume: Philosopher of moral science.* Oxford, UK: Basic Blackwell.

Fodor, J.A. (2003). *Hume variations.* Oxford, UK: Oxford University Press.

Fogelin, R.J. (1985). *Hume's scepticism in the Treatise of Human Nature.* London, England: Routledge & Kegan Paul.

Forstrom, K. J. (2010). *John Locke and personal identity: Immortality and bodily resurrection in seventeenth-century philosophy.* London, England: Continuum International Publishing Group.

Foster, J. (1991). *The Immaterial Self: A defense of the Cartesian dualist conception of the mind*. London: Routledge.

Frasca-Spada, M. (1998). *Space and the self in Hume's Treatise.* Cambridge, UK: Cambridge University Press.

Frege, G. (1956). 'The thought: A logical inquiry'. In Klemke, E.D. (Ed.). *Essays on Frege* Chicago, IL: University of Illinois Press, (1968) *Appendix A* Reprinted from *Mind*, LXV (1956) 289-311.

Frege, G. (1884/1968). *Die grundlagen der arithmetik/The foundations of arithmetic.* Translated by J. L. Austin. Evanston, IL: Northwestern University Press.

French, P.A. and Wettstein, H.K. (Eds.), and Fischer, J.M. (Guest Ed.). (2005). *Free will and moral responsibility* (Vol. No. XXIX). Oxford, UK: Blackwell Publishing Inc.

John Locke (1894). *An Essay concerning Human Understanding*. ALEXANDER CAMPBELL FRASER (Ed.). In two volumes. Oxford: Clarendon Press.

Frondizi, R. (1953). *The nature of the self: A functional interpretation.* Carbondale, IL: Southern Illinois University Press.

Fuchs, T. (2007). 'Fragmented selves: Temporality andidentity in borderline personality disorder.' Psychopathology, 40:379–387.

Gallie, I. (1936). 'Is the Self a Substance?' *Mind,* New Series, Vol. 45, No. 177, pp. 28-44.

Gallois, A. (1996). *The world without, the mind within: An essay on first-person authority.* Cambridge, UK: Cambridge University Press.

Gallois, A. (1998). *Occasions of identity: a study in the metaphysics of persistence, change, and sameness.*

Gallois, A. (2005/2011). 'Identity over time.' *Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2012/entries/identity-time/>.

Gallagher, S. (2000). 'Philosophical Conceptions of the Self: Implications for Cognitive Science'. *Trends in Cognitive Science*, Vol. 4, No. 1, pp. 14-21.

Gallagher, S. and J. Shearer (Eds.). (1999). *Models of the self*. Imprint Academic.

Gallagher S. (2011) (Ed.). *Oxford handbook of the self*. Oxford: Oxford University Press.

Ganeri, J. (2012). *The self: Naturalism, consciousness, and the first-person stance.* Oxford, UK: Oxford University Press.

Gardner, S. (1999). *Routledge philosophy guidebook to Kant and the critique of pure reason*. London, England: Routledge.

Garrett, B. (1998). *Personal identity and self-consciousness.* London, England: Routledge.

Garrett, D. (1981). 'Hume's self-doubts about personal identity.' *The Philosophical Review,* XC, No. 3, pp. 337-358.

Garrett, G. (1991). 'Vague Identity and Vague Objects.' *Noûs,* Vol. 25, No. 3, pp. 341-351.

Gasser, G. and Stefan, M. (Eds.) (2012). *Personal identity: Complex or simple?* Cambridge, UK: Cambridge University Press.

Geach, P. and Black, M. (Eds.). (1952). *Translations from the philosophical writings of Gottlob Frege.* Oxford, UK: Basil Blackwell.

Geach, P.T. (1980). *Reference and generality: An examination of some medieval and modern theories.* Cornell University Press.

Gendler, T.S. (2002). 'Personal identity and thought-experiments.' *The Philosophical Quarterly,* Vol. 52, No. 206, pp. 34-54.

Gendler, T.S. and Hawthorne, J. (2002). *Conceivability and possibility.* Oxford, UK: Clarendon Press.

Gendler, T.S. (2007).' Philosophical thought experiments, intuitions, and cognitive equilibrium.' *Midwest Studies in Philosophy*, vol. 31, issue 1, pp. 68-89.

Gettier, E. (1963). 'Is justified true belief knowledge?' Analysis, 23 (6), pp. 121-123.

Gergen, K. J. (1977). *The social construction of self-knowledge*. In Theodore Mischel (Ed.), *The Self: Psychological and Philosophical Issues.* Rowman & Littlefield.

Geach, P. (1980). *Reference and Generality*, 3rd edition. Ithaca, NY: Cornell University Press.

Gibbard, A. (1975). 'Contingent Identity,' *Journal of Philosophical Logic,* Vol. 4, No. 2, pp. 187-221.

Gilead, A. (2008). 'A Humean Argument for Personal Identity.' *Int Ontology Metaphysics*, 9, pp. 1–16.

Gilbert, R. (1949). *The concept of mind*. London: Hutchinson.

Gill, C. (Ed.) (1990). *The person and the human mind: Issues in ancient and modern philosophy.* Oxford, UK: Clarendon Press.

Gill, M.L. (1989). *Aristotle on substance: The paradox of unity.* Princeton, NJ: Princeton University Press.

Gillett, G. (1986). 'Brain Bisection and Personal Identity.' *Mind,* New Series, Vol. 95, No. 378, pp. 224-229.

Gillett, G. (1987). 'Reasoning about persons.' In Arthur R. Peacocke & Grant R. Gillett (Eds.), *Persons and personality: A contemporary inquiry.* Blackwell.

Gibb, S., Lowe, E.J. and Ingthorsson, R.D. (2013). *Mental causation and ontology*. Oxford: Oxford University Press.

Glock, H.J. (1996). *A Wittgenstein dictionary.* Oxford, UK: Blackwell Publishers.

Glover, J. (1988). *The philosophy and psychology of personal identity.* London, England: The Penguin Press.

Goldie, P. (2004). 'Emotion, Feeling, and Knowledge of the World.' In Robert C. Solomon (Ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions*. Oxford, UK: Oxford University Press.

Goldman, A. (2007). 'Philosophical intuitions: Their target, their source, and their epistemic status.' Grazer Philosophische Studien, 74 (1), pp. 1-26.

Grayling, A.C. (Ed.) (1995). *Philosophy 1: A guide through the subject Vol. 1.* Oxford, UK: Oxford University Press.

Greene, B. (1999, 2003). *The elegant universe: Superstrings, hidden dimensions, and the quest for the ultimate theory.* New York, NY: Vintage Books.

Griffin, N. J. (1977). *Relative identity.* Oxford, UK: Oxford University Press.

Grossman, R. (1973). *Ontological reduction.* Bloomington, IN: Indiana University Press.

Grossman, R. (1983). *The categorical structure of the world.* Bloomington, IN: Indiana University Press.

Grossman, R. (1965). *The Structure of Mind*. Madison and Milwaukee: The University of Wisconsin Press.

Grandin, T. and Panek, R. (2013). *The Autistic Brain: Helping Different Kinds of Minds Succeed*. New York: Mariner Books.

Hales, S.D. (1999). *Metaphysics: Contemporary readings. (1st Edition).* Belmont, CA: Wadsworth Publishing Company.

Haggvist, S. (1996). *Thought experiments in philosophy*. Almqvist & Wiksell Internat.

Hamilton, A. (2013). *The self in question: Memory, the body and self-consciousness.* London, England: Palgrave Macmillan.

Hampshire, S. et.al (1963). Edited by Pears, D.F. *David Hume: A symposium.* London, England, Macmillan & Co. Ltd.

Harris, H. (1995). *Identity: Essays based on Herbert Spencer lectures given in the University of Oxford.* Oxford, UK: Oxford University Press.

Hart, W.D. (1988). *The engines of the soul.* Cambridge, UK: Cambridge University Press.

Hasker, W. (1999). *The emergent self.* Ithaca, NY: Cornell University Press.

Haslanger, S., and Kurtz, R.M. (2006). *Persistence: contemporary readings*. Cambridge, MA: MIT Press.

Hawley, K. (2001). *How things persist.* Oxford, United Kingdom: Clarendon.

Hawley, K. (2010). 'Temporal parts.' In *Stanford Encyclopedia of Philosophy,* Winter 2010 edition. Zalta, E.N. (Ed.). [Online] Available from: http://plato.stanford.edu/archives/win2010/entries/temporal-parts/.

Hawthorne, J. (2006). *Metaphysical essays.* Oxford, UK: Oxford University Press.

Hawthorne, J. (2003). 'Identity.' In Loux, M. J., and Zimmerman, D. W. (Eds.). *The Oxford handbook of metaphysics*. Oxford, England: Oxford University Press.

Hacking, I. (1995). *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton University Press.

Heil, J. (1998). Philosophy of mind: A contemporary introduction. London, UK: Routledge.

Heil, J. (2012). *The universe as we find it*. Oxford, UK: Oxford University Press.

Heller, M. (1990). *The ontology of physical objects: Four-dimensional hunks of matter.* Cambridge, UK: Cambridge University Press.

Herbert, F. (1958). ''The mental' and the 'physical'', *Minnesota Studies in the Philosophy of Science* 2, pp. 370-497.

Hirsch, E. (1982). *The concept of identity.* New York, NY: Oxford University Press.

Hirsch, E. (1991). 'Divided Minds.' *The Philosophical Review,* Vol. 100, No. 1, pp. 3-30.

Hobson, R. P. (1993). *Autism and the development of mind.* East Essex, UK: Psychology Press, Publishers.

Hobson, P. (2011). 'Autism and the self.' In Shaun Gallagher (Ed.), *The Oxford Handbook of the Self.* Oxford, UK: Oxford University Press.

Hodges, M.P. (1990). *Transcendence and Wittgenstein's Tractatus.* Philadelphia: Temple University Press.

Hochberg, H. (1964). *Things and qualities.* Capitan and Merrill.

Hofweber, T. (2009). 'Ambitious, yet modest, metaphysics.' In David John Chalmers, David Manley & Ryan Wasserman (Eds.), *Metametaphysics: New essays on the foundations of ontology.* Oxford University Press, pp. 260—289.

Hoffman, J. and Rosenkrantz, G.S. (1994). *Substance among other categories.* Cambridge, UK: Cambridge University Press.

Hoffman, J. and Rosenkrantz, G.S. (1997). *Substance: Its nature and existence.* London, England: Routledge.

Hofstadter, D.R. and Dennett, D. (1981). *The mind's I fantasies and reflections on self and soul.* London, England: Penguin Books.

Horowitz, T., and Massey, G. (1991). *Thought experiments in science and philosophy.*
Hobbes, T., Curley E. M. (Ed.) (1994). *Leviathan*. Hackett Pub.

Hudson, H. (2012). 'The morphing block and diachronic personal identity.' In Gasser, G. and Hughes, A.L. (2012). 'The Folly of Scientism.' *The New Atlantis*, Number 37, pp. 32-50.

Hume, D. (1975). *Enquiries: Concerning human understanding and concerning the principles of morals. (3rd Edition).* Oxford, UK: Oxford University Press.

Hume, D. (1995). *British philosophy 1600-1900. David Hume. A Letter from a Gentleman.* Charlottesville, Va.: InteLex Corporation.

Hume, D. (1964). A treatise on human nature. Introduction by A.D. Lindsay. London : Dent; New York : Dutton.

Humphreys, P. (1997). 'How properties emerge', *Philosophy of Science*, 64, pp. 1-17.

Ismael, J. T. (2007). *The situated self*. Oxford: Oxford University Press.

Irwin, T. (1988). Aristotle's First Principles. Oxford: Clarendon Press.

Jackson, F. (1998). *From metaphysics to ethics.* Oxford: Oxford University Press.

Jackson, F. (2005). 'Consciousness.' In Michael Smith (Ed.). *The Oxford Handbook of Contemporary Philosophy,* Oxford: Oxford University Press, pp. 310-333.

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

Johnston, M. and Forbes, G. (1987). 'Is There a Problem about Persistence?' *Proceedings of the Aristotelian Society,* Supplementary Volumes, Vol. 61, pp. 107-135, 137-155.

Jones, D. G. (2010). 'Peering into People's Brains'. *Journal of American Scientific Affiliation*, 62, no. 2, pp. 122-32.

Jubien, M. (1997). *Contemporary metaphysics: An introduction.* Oxford, UK: Blackwell Publishers.

Kane, R. (Ed.). (2002). *The Oxford handbook of free will.* New York, NY: Oxford University Press.

Kanzian, C. (2012). 'Is "person" a sortal term?' In Gasser, G. and Stefan, M. (2012). *Personal identity: Complex or simple?* Cambridge, UK: Cambridge University Press.

Kanzian, C. (Ed.) (2008). Persistence. Ontos.

Kant, I. (1996). *Critique of pure reason (Unified Edition).* Translated by Pluhar, W.S. Cambridge, UK: Hackett Publishing Company, Inc.

Keefe, R. (2000). *Theories of vagueness.* Cambridge, UK: Cambridge University Press.

Kehr, M.W. (1916). 'The Doctrine of the Self in St. Augustine and in Descartes.' *The Philosophical Review,* Vol. 25, No. 4, pp. 587-615.

Kelly, E.F. et al. (2007). *Irreducible mind: Toward a psychology for the 21st century.* Lanham, MD: Rowman & Littlefield Publishers, Inc.

Kenny, A. (1989). *The metaphysics of mind*. Oxford, UK: Clarendon Press.

Kim, J. (1993). *Supervenience and mind: Selected philosophical essays.* New York, NY: Cambridge University Press, pp. 310-326.

Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation.* Cambridge, MA: The MIT Press.

Kim, J. (2003). *Mind in a physical world*. Cambridge: MIT Press.

Kim, J. (2005). *Physicalism, or something near enough.* Princeton, NJ: Princeton University Press.

Kim, J. (2006). *Philosophy of mind (2ⁿᵈ edition).* Cambridge, MA: Westview Press.

King, L. (1829). *Life and letters of John Locke.* London: Mdccclviii John Childs and Son Printers.

Kirk, G.S., Raven, J. and Schofield, M. (1957). *The presocratic philosophers. A critical history with a selection of texts*. Cambridge: Cambridge University Press.

Kitcher, P. (1990). *Kant's transcendental psychology.* New York, NY: Oxford University Press.

Kolak, D. and Martin, R. (1991). *Self and identity: Contemporary philosophical issues.* New York, NY: Macmillan Publishing Company.

Korfmacher, C. (2006). 'Personal identity.' *Internet Encyclopedia of Philosophy.* http://www.iep.utm.edu/person-i/

Kosso, P. (1998). *Appearance and reality: An introduction to the philosophy of physics.* Oxford, UK: Oxford University Press.

Kripke, S.A. (1972, 1980). *Naming and necessity.* Oxford, UK: Blackwell Publishing.

Kripke. S. A. (1982). *Wiggenstein on rules and private language: An elementary exposition*. Cambridge, Massachusetts: Harvard University Press.

Kripke. S. A. *Philosophical troubles*: Vol. 1. Oxford: Oxford University Press.

Kusch, Martin, 'Psychologism.' (2014). [Online] *The Stanford Encyclopedia of Philosophy* (Originally published in 2007, revised in 2011), Edward N. Zalta (Ed.), Available from: http://plato.stanford.edu/archives/spr2014/entries/psychologism/.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* Basic books.

Ladyman, J. and Don, R. (2007). *Everything must go*. Oxford: Oxford University Press.

Ladyman, J. Don, R. and Kincaid, H. (2013). *Scientific Metaphysics*. Oxford: Oxford University Press.

LaRock, E. (2008). 'Is consciousness really a brain process?' *International Philosophical Quarterly,* Vol. 48, pp. 201-229.

Laurence, B. (2010). *Epistemology: Classic problems and contemporary responses*. New York: Rowman & Littlefield Pub.

Lee, H. (1702). *Anti-Scepticism: or, Notes upon each Chapter of Mr. Locke's Essay Concerning Humane Understanding. With an Explication of all the Particulars of which He Treats, and in the same Order. In Four Books.* London: Printed for Clavel and C. Harper.

Leibniz, G.W. (1996). *New Essays on Human Understanding,* Jonathan Bennett and Peter Remnant (Eds. and Trans.), Cambridge: Cambridge University Press.

Levine, J. (1983). 'Materialism and qualia.' *Pacific Philosophical Quarterly,* 64, pp. 354-361.

Lewis, D. (1971). 'Counterparts of persons and their bodies.' *The Journal of Philosophy,* Vol. 68, No. 7, pp. 203-211.

Lewis, D. (1986). *On the plurality of words.* Oxford, UK: Blackwell Publishing.

Lewis, D. (1983). *Philosophical Papers*. Vol. 1. New York: Oxford University Press.

Lewis, D. (1988). 'Rearrangement of particles: Reply to Lowe.' *Analysis*, 48 (2), pp. 65-72.

Lewis, F. A., and Bolton, R. (1996). *Form, matter, and mixture in Aristotle.* Oxford, UK: Blackwell Publishers.

Lewis, H. D. (1978). *Persons and life after death: Essays*. Barnes & Noble.

Le Poidevin, R., Simons, P., McGonigal, A. and Cameron, R., (Eds.) (2009). *The routledge companion to metaphysics*. London: Routledge.

Gazzangia, M.S., LeDoux, J. E., & Wilson, D.H. (1977). *Language, praxis, and the right hemisphere: Clues to some mechanisms of consciousness*. Neurology, 27, 1144-1147.

Lock, D. (1971). *Memory.* London, England: The Macmillan Press Ltd.

Lockwood, M. (1989). *Mind, brain and the quantum: The compound 'I.'* Oxford, UK: Basil Blackwell.

Loewer, B in Savellos, E. E., & Yalçin, Ü. D. (1995). *Supervenience*. New York: Cambridge University Press.

LoLordo, A. (2012). *Locke's moral man*. Oxford, UK: Oxford University Press.

Loux, M.J. (1979). 'Form, species and predication in Metaphysics Z, H, and Θ.' *Mind,* 88 (349), pp. 1-23.

Loux, M.J. (1991). *Primary ousia: An essay on Aristotle's Metaphysics Z and H.* Ithaca, NY: Cornell University Press.

Loux, M.J. (1998). *Metaphysics: A contemporary introduction. (2ⁿᵈ Edition).* London, England: Routledge.

Loux, M.J. and Zimmerman, D.W. (Eds.). (2003). *The Oxford handbook of metaphysics.* Oxford, UK: Oxford University Press.

Lovibond, S. and Williams, S.G. (Eds.) (1996). *Essays for David Wiggins: Identity, truth, value IX.* Oxford/Cambridge, MA: Blackwell.

Lowe, E.J. (1987). 'Lewis on Perdurance versus Endurance.' *Analysis,* Vol. 47, No. 3, pp. 152-154.

Lowe, E.J. (1988). 'The Problems of Intrinsic Change: Rejoinder to Lewis.' *Analysis,* Vol. 48, No. 2, pp. 72-77.

Lowe, E.J. (1989a). 'Kinds of being: A study of individuation, identity and the logic of sortal terms.' *Aristotelian Society Series, Vol. 10.* Oxford, UK: Basil Blackwell.

Lowe, E.J. (1989b). 'What is a Criterion of Identity?' *The Philosophical Quarterly,* Vol. 39, No. 154, pp. 1-21.

Lowe, E.J. (1994). 'Vague Identity and Quantum Indeterminacy.' *Analysis,* Vol. 54, No. 2, pp. 110-114.

Lowe, E.J. (1995). *Locke on human understanding.* Oxon, UK: Routledge.

Lowe, E.J. (1996). *Subjects of Experience*. Cambridge: Cambridge University Press.

Lowe, E.J. (1998). *The possibility of metaphysics: Substance, identity, and time.* Oxford, UK: Oxford University Press.

Lowe, E.J. (1999). *Introduction to substance, in Metaphysics Contemporary Readings, First Edition,* in Steven D. Hales, Belmont: Wadsworth Publishing Company.

Lowe, E.J. (2000). *An introduction to the philosophy of mind.* Cambridge, UK: Cambridge University Press.

Lowe, E.J. (2000). 'Locke, Martin and substance.' *The Philosophical Quarterly,* Vol. 50, No. 201, pp. 499-514.

Lowe, E.J. (2002). *A survey of metaphysics.* Oxford, UK: Oxford University Press.

Lowe, E.J. (2006). 'Non-Cartesian substance dualism and the problem of mental causation.' *Erkenntnis,* pp. 65:5–23.

Lowe, E.J. (2008a). *Personal agency: The metaphysics of mind and action.* Oxford, UK: Oxford University Press.

Lowe, E.J. (2008b). 'Two Notions of Being: Entity and Essence.' *Royal Institute of Philosophy Supplement,* 62, pp 23-48.

Lowe, E.J. (2010). [Online] 'Ontological Dependence.' *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.). Available from: http://plato.stanford.edu/archives/spr2010/entries/dependence-ontological/.

Lowe, E.J. and Noonan, H. (1988). 'Substance, identity and time.' *Proceedings of the Aristotelian Society, Supplementary Volumes,* Vol. 62, pp. 61-100.

Loeb, L. E. (1992). *Hume Studies*, 18 (2), pp. 219-231.

Lu, M. (2013). 'The ontogenesis of the human person: A neo-Aristotelian view.' *University of St. Thomas Journal of Law and Public Policy,* 8 (1), 96-116.

Lucy, N. (2004). *A Derrida dictionary.* Oxford, UK: Blackwell Publishing.

Ludlow, P., Nagasawa, Y., and Stoljar, D. (Eds.). (2004). *There's something about Mary: essays on phenomenal consciousness and Frank Jackson's knowledge argument*. London, England: The MIT Press.

Ludwig, A. M. (1997). *How do we know who we are?.* Oxford: Oxford University Press.

Luper, S. (Ed.) (2014). 'Life's Meaning': *Cambridge Companion to Life and Death*. Cambridge: Cambridge University Press.

Lycan, W.G. (2000). *Philosophy of language: A contemporary introduction.* London, England: Routledge.

Lycan, W.G. 'Giving dualism its due.' (http://www.unc.edu/ujanel/Du.htm, accessed: 2/6/2008).

Mackie, D. (1999). 'Animalism versus Lockeanism: No contest.' *The Philosophical Quarterly,* Vol. 49, No. 196, pp. 369-376.

Mackie, J.L. (1976). *Problems from Locke.* Oxford, UK: Clarendon Press.

Makin, S. (2006). *Aristotle: Metaphysics theta*. Trans., with an introduction and commentary. Oxford: Clarendon Press.

Martin, C. B. (2008). *The mind in nature*. Oxford: Oxford University Press.

Martin, C. B. (1980). 'Substance substantiated'. *Australasian Journal of Philosophy*, 58, pp. 3–10.

Martin, C. B. (1994). 'Dispositions and Conditionals.' Philosophical Quarterly, 44(174), pp. 1-8.

Maltby, J., Day, L. and Macaskill A. (2007). *Personality, individual differences and intelligence*. Harlow: Pearson Education Limited.

Madell, G.C. (1981). *The identity of the self.* Edinburgh, Scotland: Edinburgh University Press.

Marenbon, J. (Ed.). (2009). *The Cambridge companion to Boethius*. Cambridge University Press.

Marsh, R. C. (Ed) (1956). *Betrand Russell: Logcic and knowledge, essays 1901-1950*. London: Unwin Hyman.

Martin, R. and Barresi, J. (2000). *Naturalization of the soul: Self and personal identity in the eighteen century.* London, England: Routledge.

Martin, R. and Barresi, J. (Eds.) (2003). *Personal identity.* Malden, MA: Blackwell Publishing.

Martin, R. and Barresi, J. (2006*). The rise and fall of soul and self: An intellectual history of personal identity.* New York, NY: Columbia University Press.

Martin, R. (1998). *Self-concern: An experiential approach to what matters in survival*. Cambridge: Cambridge University Press.

Matthews, G.B. (1967). 'The Inner Man.' *American Philosophical Quarterly,* Vol. 4, No. 2, pp. 166-172.

Maudlin, T. (1994, 2002). *Quantum non-locality and relativity: Metaphysical intimations of modern physics*. (2nd Edition). Oxford, UK: Blackwell Publishing.

Mauss, M. (1985). 'A category of the human mind: The notion of person; the notion of self.' (W.D. Halls, Trans.). In M. Carrithers, S. Collins, and S. Lukes (Eds.), *The category of the person: Anthropology, philosophy, history.* Cambridge, England: Cambridge University Press, pp. 1-25. (Original work published 1938)

Malpas, J. and Solomon, R. (1998). *Death and Philosophy*. London: Routledge.

McCann, E. (2001). 'Locke's theory of substance under attack!' *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition,* Vol. 106, No. ½, pp. 87-105.

McCall, G. J. (1977). 'The social looking-glass: A sociological perspective on self-development.' In Theodore Mischel (Ed.), *The Self: Psychological and Philosophical Issues.* Rowman & Littlefield.

McCall, C. (1990). *Concepts of person: An analysis of concepts of person, self and human being.* Sydney, Australia: Avebury.

McGinn, C. (1991/1993). *The Problem of consciousness: essays toward a resolution*. Oxford: Basil Blackwell.

McGinn, M. (1997). *Wittgenstein and the Philosophical Investigations.* London, England: Routledge.

McGinn, C. (1999). *The mysterious flame: Conscious minds in a material world*. New York: Basic Books.

McGinn, C. (1999/2004). *Consciousness and its Object*. Oxford: Clarendon; New York: Oxford University Press.

McGinn, C. (2012). *Truth by analysis: Games, names and philosophy.* New York, NY: Oxford University Press.

McTaggert, J.M.E. (1921). *The nature of existence., Vol. 1.* Cambridge: Cambridge University Press.

MacIntyre, A. (Ed.). *Hume's Ethical Writings: Selections from David Hume*, Collier Books / Macmillan 1965; University of Notre Dame Press, 1979.

Menn, S. (1998). *Descartes and Augustine.* Cambridge, MA: Cambridge University Press.

Merleau-Ponty, Maurice (1962). *The phenomenology of perception.* Colin Smith (Trans.). New Jersey: Routledge and Kegan Paul.

Merricks, T. (1997). 'Fission and personal identity over time.' *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition,* Vol. 88, No. 2, pp. 163-186.

Merricks, T. (1999). 'Endurance, Psychological Continuity, and the Importance of Personal Identity.' *Philosophy and Phenomenological Research,* Vol. 59, No. 4, pp. 983-997.

Merricks, T. (2001). *Objects and persons.* Oxford, UK: Oxford University Press.

Merricks, T. (1994). 'Endurance and Indiscernibility.' *Philosophy*, 91 (4):165-184.

Metzinger, T. (2009). *The ego tunnel: The science of the mind and the myth of the self.* New York, NY: Basic Books.

Millican, P. (2002). *Reading Hume on human understanding: Essays on the first Enquiry.* Oxford, UK: Clarendon Press.

Mischel, T. (Ed.). (1977). *The self psychological and philosophical issues.* Oxford, UK: Basil Blackwell.

Montefiore, A. (1973). *Philosophy and personal relations.*

Moore, G.E. (1959). *A defense of common sense in philosophical papers.* London: George Allen & Unwin. Reprinted in Louis Pojman (2001). *Vol. III The Twentieth Century.* New York, NY: Oxford University Press.

Moreland, J.P. (2001). *Universals.* London, UK: McGill-Queen's University Press.

Moreland, J.P. (2008). *Consciousness and the existence of God: A theistic argument.* New York, NY: Routledge.

Moreland, J.P. (2009). *The recalcitrant imago Dei: Human persons and the failure of naturalism*. London: SCM Press.

Moreland, J.P. and Rae, S.B. (2000). *Body & soul: Human nature and the crisis in ethics*. Downers Grove, IL: InterVarsity Press.

Montefiore, A. (Ed.) (1973). *Philosophy and Personal Relations: an Anglo-French Study*.
Moris, W.E and Brown, C. R. (2014). 'David Hume'. Encyclopedia of Philosophy, ed. E.N. Zalta http://plato.stanford.edu/archives/sum2014/entries/hume.

Munitz, M.K. (Ed) (1971). *Identity and individuation.* New York, NY: New York University Press.

Murphy N., and Brown, W.S. (2007). *Did my neurons make me do it? Philosophical and neurobiological perspectives on moral responsibility and free will.* Oxford, UK: Oxford University Press.

Nagel, T. (1974). 'What it is like to be a bat', *The Philosophical Review* 83: 4, pp. 435-450.

Nagel, T. (1975). *Brain bisection and the unity of consciousness.* In *Personal Identity*. Ed. John Perry. Los Angeles, CA: University of California Press.

Nidditch, P.H. (Ed.). (1975). *John Locke: An essay concerning human understanding.* Oxford, UK: Clarendon Press.

Ninan, D. (2009). 'Persistence and the First-Person Perspective.' *Philosophical Review,* Vol. 118, No. 4, pp. 425-464.

Noonan, H. (1989, 1991, 2003). *Personal identity. (2nd Edition).* London, England: Routledge.

Noonan, H. (1991). 'Indeterminate Identity, Contingent Identity and Abelardian Predicates.' *The Philosophical Quarterly,* Vol. 41, No. 163, pp. 183-193.

Noonan, H. (Ed.) (1993). 'Introduction.' In *Identity*. Sydney, Australia: Dartmouth Publishing Co. Ltd., pp. xi-xx.

Noonan, H. (1998). 'Animalism versus Lockeanism: A current controversy.' *The Philosophical Quarterly,* Vol. 48, No. 192, pp. 302-318.

Noonan, H. (1999). *Hume on knowledge.* London, England: Routledge.

Noonan, H. (2001). 'Animalism versus Lockeanism: Reply to Mackie.' *The Philosophical Quarterly,* Vol. 51, No. 202, pp. 83-90.

Noonan, H. and Curtis, B. (2014). [Online] 'Identity.' *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.). Available from: http://plato.stanford.edu/archives/sum2014/entries/identity/.

Noonan, H. W. (1980). *Objects and identity*. The Hague: Martinus Nijhoff.

Norden, J. (2007). *Understanding the brain.* Chantilly, VA: The Teaching Company.

Norton, D.F. (1982). *David Hume: Common-sense moralist, sceptical metaphysician.* Princeton, NJ: Princeton University Press.

Norton, D.F. (Ed.). (1993). *The Cambridge companion to Hume.* Cambridge, UK: Cambridge University Press.

Norton, D. F. and Norton, M. J. (2000). *A treatise of human nature: Being an attempt to introduce the experimental method of reasoning into moral subjects. Oxford: Oxford University Press.*

Nozick, R. (1981). *Philosophical explanations.* Oxford, UK: Clarendon Press.

Novotný, D. D., & Novák, L. (Eds.) (2014). *Neo-Aristotelian perspectives in metaphysics*. London: Routledge.

O'Brien, D. (2006). *An introduction to the theory of knowledge.* Cambridge, UK: Polity Press.

O'Brien, L. (2007). *Self-knowing agents.* Oxford, UK: Oxford University Press.

O'Connor, T. and Jacobs, J.D. (2003). 'Emergent Individuals.' *The Philosophical Quarterly,* Vol. 53, No. 213, pp. 540-555.

O'Connor, T. and Wong, Yu H. (2005). 'The metaphysics of emergence', *Noûs* 39:4, pp. 658-678.

O'Connor, T. and Wong, Yu H. (2002/2012). [Online] 'Emergent properties', in Edward N. Zalta (ed.), Stanford Encyclopaedia of Philosophy (2002 Edition rev., 2012) Available from: http://plato.stanford.edu/entries/properties-emergent/.

Oderberg, D.S. (1993). *The metaphysics of identity over time.* London, England: The Macmillan Press Ltd.

Oderberg, D.S. (Ed) (1999). *Form and matter: Themes in contemporary metaphysics.* Oxford, UK: Blackwell Publishers.

Oderberg, D.S. (2000). *Abortion in Applied Ethics: A non-consequentialist approach.* Oxford, UK: Blackwell Publishers.

Oderberg, D.S. (2007). *Real Essentialism*. London: Routledge.

Oderberg, D. S. (2008). 'The metaphysical status of the embryo: Some arguments revisited.' *Journal of Applied Philosophy*, 25 (4), pp. 263-276.

Oderberg, D. S. (2012). 'Hume, the occult, and the substance of the school.' *Metaphysica*, *13*(2), 155-174.

O'Hear, A. (1985). *What philosophy is: An introduction to contemporary philosophy.* London: Penguin Books.

Olson, E.T. (1997). *The human animal: Personal identity without psychology.* Oxford, UK: Oxford University Press.

Olson, E. T. (1998). 'There is no problem of the self.' *Journal of Consciousness Studies*, *5*(5-6), pp.645-657.

Olson, E.T. (2002). 'What does functionalism tell us about personal identity?' *Nous,* 36 (4). pp. 682-698.

Olson, E.T. (2007). *What are we? A study of personal ontology.* Oxford, UK: Oxford University Press.

Olson, E.T. (2010). 'Personal Identity.' *The Stanford Encyclopedia of Philosophy,* Edward N. Zalta (Ed.), URL = <http://plato.stanford.edu/archives/win2010/entries/identity-personal/>.

Owens, J. (1988). 'The Self in Aristotle.' *The Review of Metaphysics,* Vol. 41, No. 4, pp. 707-722.

Papineau, D. (1993). *Philosophical Naturalism.* Oxford: Blackwell.

Patricia, C. S. (1986). *Neurophilosophy: Toward a unified science of the mind-brain*. Massachu.: MIT Press.

Parfit, D. (1971). 'Personal identity.' *The Philosophical Review,* Vol. 80, No. 1, pp. 3-27.

Parfit, D. (1971). 'The importance of self-identity.' *The Journal of Philosophy,* Vol. 68, No. 20, pp. 683-690.

Parfit, D. (1984). *Reasons and persons.* Oxford, UK: Clarendon Press.

Parfit, D. (2012). 'We Are Not Human Beings.' *Royal Institute of Philosophy, 87,* pp 5-28.

Parnas, J., and Sass, L. (2011). 'The structure of self-consciousness in schizophrenia.' In Shaun Gallagher (Ed.), *The Oxford Handbook of the Self.* Oxford, UK: Oxford University Press.

Parsons, T. (2000). *Indeterminate identity: Metaphysics and semantics.* Oxford, UK: Clarendon Press.

Pasnau, R. (2011). *Metaphysical themes 1274-1671*. Oxford, UK: Oxford University Press.
Paul, L. (1961). *Persons and perception.* London: Faber and Faber.

Peacocke, A. and Gillett, G. (1987). *Persons and personality: A contemporary inquiry.* Oxford, UK: Basic Blackwell.

Pears, D. (1990). *Hume's system: An examination of the first book of his Treatise.* Oxford, UK: Oxford University Press.

Peijnenburg, J. and Atkinson, D. (2003). 'When are thought experiments poor ones?' *Journal for General Philosophy of Science 34:* 305–322, 2003.

Penelhum, T. (1975). *Hume.* London, England: The Macmillan Press Ltd.

Penelhum, T. (1992). *David Hume: An introduction to his philosophical system.* West Lafayette, IN: Purdue University Press.

Penelhum, T. (2000/2003). *Themes in Hume: The self, the will, religion.* Oxford, UK: Oxford University Press.

Perry, J. (1972). 'Can the Self Divide?' *The Journal of Philosophy,* Vol. 69, No. 16, pp. 463-488.
Perry, J. (Ed.). (1975). *Personal identity.* Berkeley, CA: University of California Press.

Perry, J. (1976). *The importance of being identical, In Amelie Oksenberg Rorty (Ed.), The identities of persons.* University of California Press, pp. 67-90.

Perry, J. (1977). *A dialogue on personal identity and Immortality.* Indianapolis, MN: Hackett Publishing Company.

Perry, J. (2002). *Identity, personal identity, and the self.* Indianapolis, IN: Hackett Publishing Company, Inc.

Peunenburg, J. and Atkinson, D. (2003). 'When are thought experiments poor ones?' *Journal for General Philosophy of Science,* 34, pp. 305-322.

Pike, N. (1967). 'Hume's bundle theory of the self: A limited defense.' *American Philosophical Quarterly,* Vol. 4, No. 2, pp. 159-165.

Pitson, A.E. (2002). *Hume's philosophy of the self.* London, England: Routledge.

Place, U.T. (1956). 'Is consciousness a brain process?' *British Journal of Psychology* 47, pp. 44-50.

Plantinga, A. (1974). *The nature of necessity.* Oxford, UK: Clarendon Press.

Plantinga, A. eby Davidson M. (2003). *Essays in the metaphysics of modality.* Oxford, UK: Oxford University Press.

Price, H.H. (1932). *Perception.* London, England: Methuen & Co. Ltd.

Price, H.H. (1940). *Hume's theory of the external world.* Oxford, UK: Clarendon Press.

Prinz, J.J. (2012). *The conscious brain*. Oxford: Oxford University Press.

Popper, K. (1963, 1994). *Conjectures and refutations: The growth of scientific knowledge.* London, England: Routledge.

Popper, K.R. and Eccles, J.C. (1977). *The self and its brain.* London, England: Springer International.

Polansky, R. (2007). *Aristotle's de anima*. Cambridge: Cambridge University Press.

Pruss, A.R. (2009). 'Artificial intelligence and personal identity.' *Faith and Philosophy: Science and Human Nature,* Vol. 26, Issue 5, pp. 487-500.

Puccetti, R. (1968). *Persons: a study of possible moral agents in the Universe*. London: Macmillan.

Putnam, H. (1977). 'Realism and Reason.' *Proceedings and Addresses of the American Philosophical Association,* Vol. 50, No. 6, pp. 483-498.

Quine, W.V. (1969). *Ontological relativity and other essays. (No. 1).* New York: Columbia University Press.

Quine, W.V. (1976). 'Worlds away.' *The Journal of Philosophy,* Vol. 73, No. 22, pp. 859-863.

Quinn, P.L. (1978). 'Personal Identity, Bodily Continuity and Resurrection.' *International Journal for Philosophy of Religion,* Vol. 9, No. 2, pp. 101-113.

Quinton, A. (1973). *The nature of things.* London: Routledge & Kegan.

Radden, J. (2011). 'Multiple selves.' In Shaun Gallagher (Ed.), *The Oxford Handbook of the Self.* Oxford, UK: Oxford University Press, pp. 547-570.

Rana, F. (2011). *Creating life in the lab: How new discoveries in synthetic biology make a case for the creator.* Grand Rapids, MI: BakerBooks.

Rae, M., (Ed.) (2009). *Oxford readings in philosophical theology: Vol. II, Providence, scripture and resurrection.* Oxford: Oxford University Press.

Rana, F. and Ross, H. (2004). *Origins of life.* Colorado Springs, CO: NavPress.

Rawls, J. (1955). 'Two Concepts of Rules.' *The Philosophical Review,* Vol. 64, No. 1, pp. 3-32.

Rawls, J. (1971, 1999). *A Theory of Justice*. Massachusetts: The Belknap Press of Harvard University Press.

Recanati, F. (2009). 'De re and De se.' *Dialectica*, Vol. 63, No. 3, pp. 249–269.

Rescher, N. (2005). *What if?: Thought experimentation in philosophy*. New Brunswick, NJ: Transaction Publishers.

Roache, R. (2006). 'A Defence of Quasi-Memory.' *Philosophy,* Vol. 81, No. 2, pp. 323-355.

Robinson, D. (1982). 'Re-Identifying Matter.' *The Philosophical Review,* Vol. 91, No. 3 pp. 317-341.

Robinson, D. (1985). 'Can amoebae divide without multiplying?' *Australasian Journal of Philosophy,* 63:3, pp. 299-319.

Robinson, D. (1989). 'Matter, motion, and Humean supervenience.' *Australasian Journal of Philosophy,* 67:4, pp. 394-409.

Robinson, H. (1994). *Perception.* London, England: Routledge.

Robinson, H. (2004). 'Thought experiments, ontology, and concept-dependent truthmakers.' *The Monist,* Vol. 87, No. 4, pp. 537-553.

Robinson, H. (2014). 'Substance.' *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/substance/>.

Rorty, A.O. (Ed.) (1976). *The identities of persons.* Berkeley, CA: University of California Press.

Rorty, A.O. (1990). 'Persons and personae.' In Gill, C. *The person and the human mind: Issues in ancient and modern philosophy.* In Ancient and Modern Philosophy. Oxford, UK: Oxford University Press.

Rorty, A. (Ed.). (1992). *The Identities of Persons.* University of California Press.

Ross, W.D. (1924). *Aristotle's metaphysics. (Vol.I).* Oxford, UK: Clarendon Press.

Ross, W.D. (1924). *Aristotle's metaphysics. (Vol.II).* Oxford, UK: Clarendon Press.

Rovane, C. (1998). *The bounds of agency: An essay in revisionary metaphysics.* Princeton, NJ: Princeton University Press.

Royle, N. (2003). *Jacques Derrida.* London, England: Routledge.

Rudd, A. (2012). *Self, value, and narrative: A Kierkegaardian approach.* Oxford, UK: Oxford University Press.

Russell, B. (1912, 1959, 1970). *The problems of philosophy.* London, England: Oxford University Press.

Rychter, P. (2009). 'There is No Puzzle about Change.' *Dialectica,* Vol. 63, No. 1, pp. 7–22.

Ryle, G. (1949, 1963, 1990). *The Concept of Mind.* London, UK: The Penguin Group.

Ryle, G. (1963). *The concept of mind*. Harmondsworth: Penguin.

Sainsbury, R.M. and Rye, M. (2012). *Seven puzzles of thought and how to solve them: An originalist theory of concepts.* Oxford, UK: Oxford University Press.

Salmon, N. U. (2005). *Metaphysics, mathematics, and meaning*. Collected papers, vol. 1. Oxford.

Savellos, E. E., & Yalçin, Ü. D. (1995). *Supervenience*. New York: Cambridge University Press.

Scaltase, T. (1994). *Substances and universals in Aristotle's metaphysics*. Cornell University Press.

Schaffer, J. (2010). 'Monism: The priority of the whole.' *Philosophical Review,* Vol. 119, No. 1, pp. 31-76.

Schechtman, M. (1996). *The Constitution of Selves*. Ithaca, New York: Cornell University Press.

Schechtman, M. (2014). *Staying alive: Personal identity, practical concerns, and the unity of a life.* Oxford: Oxford University Press.

Schlesinger, G.N. (1983). *Metaphysics: Methods and problems.* Oxford, UK: Basil Blackwell.

Schmitt, F.F. (2014). *Hume's epistemology in the Treatise.* Oxford, UK: Oxford University Press.

Schwartz, R. (2004). *Perception.* Oxford, UK: Blackwell Publishing.

Schwerin, A. (2012). 'Hume on the Self.' *Int Ontology Metaphysics,* 13, pp. 65–85.

Schwyzer, H. (1997). 'Subjectivity in Descartes and Kant.' *The Philosophical Quarterly,* Vol. 47, No. 188, pp. 342-357.

Schwartz, J.M., and Begley, S. (2002). *The Mind and the Brain: Neuroplasticity and the Power of Mental Force*. New York: HarperCollins Pub.

Searle, J.R. (1969). *Speech acts: An essay in the philosophy of language.* Cambridge, UK: Cambridge University Press.

Searle, J.R. (1992). *The rediscovery of the mind*. London, England: MIT press.

Searle, J.R. (1984). *Minds, brains and science*. Cambridge, MA: Harvard University Press.

Searle, J. R. (2004). *The Mind: A brief introduction*. Oxford: Oxford University Press.

Selby-Bigge, L.A. (Ed.). (1978). *David Hume: A treatise of human nature.* Oxford, UK: Clarendon Press.

Seung, S. (2012). *Connectome: How the brain's wiring makes us who we are*. Boston, MA: Houghton Mifflin Harcourt.

Sergeant, J. (1697). *Solid philosophy asserted against the fancies of the ideists, or the method to science farther illustrated: with reflections on Mr. Locke's essay concerning human understanding*. London: Printed for Roger Clavil…Abel Roper…and Thomas Metcalf….

Seigel, J. (2005). *The idea of the self: Thought and experience in Western Europe since the seventeenth century*. Cambridge, MA: Cambridge University Press.

Shoemaker, S. (1963). *Self-knowledge and self-identity.* Ithaca, NY: Cornell University Press.

Shoemaker, S. (1968). 'Self-reference and self-awareness.' *Journal of Philosophy,* 65 (19), pp. 555-567.

Shoemaker, S. (1970). 'Persons and their pasts.' *American Philosophical Quarterly*, *7*(4), 269-285.

Shoemaker, S. (2003). *Identity, cause and mind.* Oxford, UK: Clarendon Press.

Shoemaker, S. and Swinburne, R. (1984). *Personal identity.* Oxford, UK: Basic Blackwell.

Sider, T. (2000). 'Recent work on identity over time.' *Philosophical Books,* 41, pp. 81-89.
Sider, T. (2001a). *Four-dimensionalism*. Oxford: Clarendon Press.

Sider, T. (2001b). 'Criteria of personal identity and the limits of conceptual analysis.' *Philosophical Perspectives,* 15, pp. 189–209.

Sider, T. (2012). *Writing the Book of the World* (2011).Oxford: Oxford University Press.

Sider, T. Hawthorne, J. and Zimmerman, D.W. (Eds.) (2007). *Contemporary debates in metaphysics.* Wiley-Blackwell.

Siderits, M., Thompson, E. and Zahavi, D. (Eds.) (2011). *Self, no self? Perspectives from analytical, phenomenological, and Indian traditions.* Oxford, UK: Oxford University Press.

Simons, P. (1994). 'Particulars in particular clothing: Three trope theories of substance.' *Philosophy and Phenomenological Research*, pp. 553-575.

Simons, P.M. (1998). *Farewell to substance: A differentiated leave-taking.* Oxford, UK: Blackwell Publishers Ltd.

Singer, P. (Ed.) (2006). *In defense of animals: The second wave.* Oxford, UK: Blackwell Publishing.

Singer, P. (1993). *Practical Ethics*. Cambridge: Cambridge University Press.

Sidelle, A. (1989). *Necessity, Essence, and Individuation: A Defense of Conventionalism.* Ithaca, N. Y. Cornell University Press.

Smart, J.J.C. (1959). 'Sensations and brain processes', *Philosophical Review* 68, pp. 141-156.

Smith, N. K. (1941). *The philosophy of David Hume: a critical study of its origins and central doctrines.* Macmillan.

Smith, J.K. (2007). (Trans.). *Immanuel Kant's critique of pure reason.* London, England: The Macmillan Press Limited.

Smith, N. K. (1929). (Trans.). *Immanuel Kant's Critique of Pure Reason*. London: Macmillan Press.

Smith, Q. and Oaklander, L.N. (1995). *Time, change and freedom: An introduction to metaphysics.* London, England: Routledge.

Sorabji, R. (2006). *Self: Ancient and modern insights about individuality, life and death.* Chicago, IL: University of Chicago Press.

Sorabji, R. (2008). *Self: Ancient and modern insights about individuality, life, and death*. University of Chicago Press.

Sorensen, R. A. (1992). *Thought experiments.* Oxford, UK: Oxford University Press.

Sorensen, R. (2006). 'Meta-conceivability and thought experiments.' In *The Architecture of the Imagination: New Essays on Pretence, Possibility and Fiction.* Shaun Nichols (Ed.). Oxford, UK: Oxford University Press, pp. 257-272.

Sorensen, R. (2012). 'Vagueness.' *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2013/entries/vagueness/>.

Sosa, E. (1990). Surviving Matters. *Noûs* 24 (2), pp. 297-322.

Sperry, R.W. (1968). 'Hemisphere deconnection and unity in conscious awareness.' *American Psychologist,* Vol. 23, No. 10, pp. 723-733.

Spinoza, B. (2000). *Ethics* (GHR Parkinson, Trans.). London: Everyman's Library.

Sprague, E. (1978). *Metaphysical thinking.* New York, NY: Oxford University Press.

Stanley, J. (1988). 'Persons and their properties.' *The Philosophical Quarterly,* Vol. 48, No. 191, pp. 159-175.

Stead, C. (1977). *Divine Substance.* Oxford, UK: Oxford University Press.

Sutton, J. (2010). Memory. Stanford *Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/memory/>.

Davis, S. (1993). *Risen Indeed: Making Sense of the Resurrection*. Grand Rapids, Michigan: Wm. B. Eerdmans Publishing Company.

Stocker, B. (2006). *Derrida on deconstruction.* London, England: Routledge.

Stone, J. (1987). 'Why potentiality matters.' *Canadian Journal of Philosophy,* 17 (December):815-829.

Stone, J. (2005). 'Why There Still Are No People.' *Philosophy and Phenomenological Research,* Vol. 70, No. 1, pp. 174-192.

Strawson, P. F. (1959). *Individuals: An essay in descriptive metaphysics.* London, England: Methuen.

Strawson, P.F. (1966). *The bounds of sense: An essay on Kant's Critique of Pure Reason.* London, England: Routledge.

Stroud, B. (1977). *Hume*. London, UK: Routledge and Kegan Paul plc.

Stroud, B. (1984). *The significance of philosophical skepticism*. Oxford: Clarendon Press.

Swinburne, R. (1968, 1981). *Space and time. (2nd Edition).* Hong Kong: The Macmillan Press Ltd.

Swinburne, R. (1997). *The Evolution of the Soul,* rev. ed. Oxford, UK: Clarendon Press.

Tahko, T. E. (Ed.) (2012). *Contemporary Aristotelian Metaphysics*. Cambridge University Press.

Taylor, R. (1974, 1963). *Metaphysics (2nd Edition).* Eaglewood Cliffs, NJ: Prentice-Hall, Inc.
Taylor, C. (1989). *Sources of the self: The making of the modern identity.* Cambridge, MA: Harvard University Press.

Teichman, J. (1985). 'The Definition of Person.' *Philosophy,* Vol. 60, No. 232, pp. 175-185.

Thiel, U. (2006). 'Self-consciousness and personal identity.' In The Cambridge History of Eighteenth-Century Philosophy, Volume 1. Cambridge, MA: Cambridge University Press.

Thomason, R.H. (1982). 'Identity and vagueness.' *Philosophical Studies*, 42, pp. 329-332.

Thomasson, A. (2013). 'Categories.' *The Stanford Encyclopedia of Philosophy.* Edward N. Zalta (Ed.), URL = <http://plato.stanford.edu/archives/fall2013/entries/categories/>.

Thomasson, A.L. (2007). *Ordinary objects.* Oxford, UK: Oxford University Press.

Thomson, J.J. (1983). 'Parthood and Identity Across Time.' *The Journal of Philosophy,* Vol. 80, No. 4, pp. 201-220.

Thomson, J.J. (2008). In Theodore Sider, John Hawthorne & Dean W. Zimmerman (eds.), *Contemporary debates in metaphysics.* Blackwell Pub.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind.* Cambridge, MA: The Belknap Press of Harvard University Press.

Thompson, M. (2009). *Me.* Stocksfield, UK: Acumen Publishing Limited.

Tipton, I.C. (Ed.). *Locke on human understanding: Selected essays.* Oxford, UK: Oxford University Press.

Toner, P. (2009). 'Personhood and Death in St. Thomas Aquinas.' *History of Philosophy Quarterly,* Vol. 26, No. 2, pp. 121-138.

Tooley, M. (1986). 'Abortion and infanticide.' In Singer, P. (Ed.). *Applied Ethics*. Oxford, UK: Oxford University Press.

Tooley, M. (1972). 'Abortion and infanticide'. *Philosophy and Public Affairs*, Vol., 2, No. 1, pp. 37-65.

Toulmin, S.E. (1977). 'Self-knowledge and knowledge of the 'self'.' In Michael, T. (Ed.) *The Self: Psychological and philosophical issues.* Oxford, UK: Basil Blackwell.

Traiger, S. (Ed.). (2006). *The Blackwell guide to Hume's Treatise.* Oxford, UK: Blackwell Publishing.

Traiger, S. (1985). 'Flage on Hume's account of memory.' *Hume's Studies*, 11 (2), pp. 166-172.

Tweyman, S. (1995). *David Hume: Critical assessments. (Vol. I).* London, England: Routledge.

Tweyman, S. (1995). *David Hume: Critical assessments. (Vol. II).* London, England: Routledge.

Tweyman, S. (1995). *David Hume: Critical assessments. (Vol. III).* London, England: Routledge.

Tweyman, S. (1995). *David Hume: Critical assessments. (Vol. IV).* London, England: Routledge.

Tye, M. (1995). *The problem of consciousness*. Cambridge: MIT Press.

Tye, M. (2003). *Consciousness and persons*. Cambridge: MIT Press, 2003.

Uebel, T. (2012). Carnap, philosophy and "politics in its broadest sense''. In R. Creath (ed.), *Rudolf Carnap and the Legacy of Logical Empiricism*. Springer Verlag. 133—148.

Unger, P. (1990). *Identity, consciousness and value.* Oxford, UK: Oxford University Press.

Unger, P. (1979). 'I do not Exist', in *Perception and Identity*, G. F. MacDonald (ed.), London: Macmillan.

Van Inwagen, P. (1983). *An essay on free will.* Oxford, UK: Clarendon Press.

Van Inwagen, P. (1990). *Material beings.* Ithaca, NY: Cornell University Press.

Van Inwagen, P. (1993). *Metaphysics.* Boulder, CO: Westview Press.

Van Inwagen, P. and Zimmerman, D. (2007). *Persons: Human and divine*. Oxford: Oxford University press.

Van Inwagen, P. (1997). 'Materialism and the Psychological-Continuity Account of Personal Identity.' *Noûs,* Vol. 31, Supplement: Philosophical Perspectives, 11, Mind, Causation, and World, pp. 305-319.

Varela, FJ, Thompson, E. and Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* Cambridge, MA: The MIT Press.

Vesey, G. (1974). *Personal identity (Problems of philosophy).* Macmillan.

Wasserman, R. (2004). 'Framing the Debate over Persistence.' In Hüntelmann, R., Meixner, U., and Tegtmeier, E. (Eds.). *Metaphysica.* Vol. 5, No. 1, pp. 67-80.

Wasserman, R. (2011). 'Personal Identity, Indeterminacy, and Obligation.' In Gasser, G. and M. Stefan, M. (Eds.). *Personal Identity: Simple or Complex?* Cambridge: Cambridge University Press.

Waxman, W. (1994). *Hume's theory of consciousness.* Cambridge, UK: Cambridge University Press.

Westerhoff, J. (2005). *Ontological categories.* Oxford, UK: Clarendon Press.

Westerhoff, J. (2011). *Reality: A very short introduction.* Oxford, UK: Oxford University Press.

Wheeler, S.C. III (2000). *Deconstruction as analytical philosophy.* Stanford, CA: Stanford University Press.

White, S.L. (1991). *The unity of the self.* Cambridge, MA: The MIT Press.

Wiggins, D. (1967). Identity and spatio-temporal continuity. Oxford: B. H. Blackwell.

Wiggins, D. (2001). *Sameness and substance renewed.* Cambridge, UK: Cambridge University Press.

Wilkes, K.V. (1978). 'Consciousness and commissurotomy.' *Philosophy,* Vol. 53, No. 204, pp. 195-199.

Wilkes, K.V. (1988). *Real people: Personal identity without thought experiments.* Oxford, UK: Clarendon Press.

Wilkerson, T.E. (1974). *Minds, brains and people.* Oxford, UK: Clarendon Press.

Williams, B. (1973). *Problems of the self: Philosophical papers 1956-1972.* Cambridge, UK: Cambridge University Press.

Williams, B. (1985). *Ethics and the Limits of Philosophy*. Harvard University Press.

Williams, C. J. F. (1989). *What is identity?* Oxford: Clarendon Press.

Williams, C.J.F (1981). *What is existence?* Oxford, UK: Oxford University Press.

Williams, D. C. (1953). 'On the elements of being: I.' *The review of metaphysics*, 3-18.

Williams, N.E. (2010). 'Putting powers back on multi-track.' Forthcoming: *Philosophia.*

Williamson, T. (1986). 'Criteria of Identity and the Axiom of Choice.' *The Journal of Philosophy,* Vol. 83, No. 7, pp. 380-394.

Williamson, T. (1990). *Identity and discrimination.* Malden, MA: Basil Blackwell.

Williamson, T. (1994): *Vagueness*. London, England: Routledge,

Williamson, T. (2007). *The philosophy of philosophy.* Oxford, UK: Blackwell Publishing.

Williams, M. (1993). *Scepticism*. Aldershot: Dartmouth.

Wippel, J. F. (2000). *The metaphysical thought of Thomas Aquinas: from finite being to uncreated being* (Vol. 1). CUA Press.

Wittgenstein, L. (1958). *The blue and brown books (1889-1951).* Oxford, UK: Blackwell.

Wittgenstein, L. (1961, 1974). *Tractatus logico-philosophicus.* Pears, D.F. and McGuinness, B.F. (Trans.). London, England: Routledge.

Wittgenstein, L. (1976). *Philosophical investigations.* Translated by Anscombe, G.E.M. Oxford, UK: Basil Blackwell.

Wittgenstein, L. (2009). *Philosophische Untersuchungen: Philosophical investigations. (Revised 4th Ed.).* Translated by Anscombe, G.E.M., Hacker, P.M.S. and Schulte, J.

Wolfram, S. (1974). 'Hume on personal identity.' *Mind,* New Series, Vol. 83, No. 332, pp. 586-593.

Wollheim, R. (1984). *The thread of life.* Cambridge, MA: Harvard University Press.

Wood, A.W. (1984). *Self and nature in Kant's philosophy.* Ithaca, NY: Cornell University Press.

Woolhouse, R.S. (1971). *Locke's philosophy of science and knowledge.* Oxford, UK: Basil Blackwell.

Woolhouse, R. (Ed.). (1988). *George Berkeley: Principles of human knowledge and three dialogues between Hylas and Philonous.* London, England: Penguin Books.

Woolhouse, R.S. (1993). *Descartes, Spinoza, Leibniz: The concept of substance in seventeenth-century metaphysics*. New York: Routledge, pp. 94-95.

Yagisawa, Takashi (2011). 'Précis of Worlds and Individuals, Possible and Otherwise.' *Analytic Philosophy*, 52 (4), pp. 270-272.

Yolton, J.W. (1985). *Locke: An introduction.* Oxford, UK: Basic Blackwell.

Yolton, J.W. (1995). *A Locke dictionary.* Oxford, UK: Blackwell Publishers.

Zemach, E.M. (1969). 'Personal identity without criteria.' *Australasian Journal of Philosophy, 47:3, 344-353.*

Zemach, E.M. (1972). 'The Reference of 'I'.' *Philosophical Studies, 23*, 68-75.

Zimmerman, D.W. (1998). 'Criteria of identity and the 'Identity Mystics'.' *Erkenntnis (1975-),* Vol. 48, No. 2/3, Analytical Ontology, pp. 281-301.

Zohar, D. (1990). *The quantum self.* London, UK: Bloomsbury Publishing PLC.