Persons—Their Identity and Individuation

Roger Melin
Roger Melin

PERSONS—THEIR IDENTITY AND INDIVIDUATION

Umeå, 1998

Institutionen för filosofi och vetenskapsteori
Umeå universitet
901 87 Umeå

Akademisk avhandling
som med vederbörligt tillstånd av rektorämbetet vid Umeå universitet
för avläggande av filosofi doktorsexamen
ABSTRACT
This study is about the nature of persons and personal identity. It belongs to a tradition that maintains that in order to understand what it is to be a person we must clarify what personal identity consists in. In this pursuit, I differentiate between the problems (i) How do persons persist? and (ii) What facts, if any, does personal identity consist in?

In chapters 2-3, I discuss matters related to the first question. In chapter 2, I discuss ‘identity’ and ‘criterion of identity’. I argue that we ought to understand ‘identity’ as numerical identity. A ‘criterion of identity’, I argue, should be understood as a specification of the essential conditions for being an object of some sort S. In chapter 3, I distinguish between two different accounts of how persons persist; the endurance view (persons persist three-dimensionally through time), and the perdurance view (persons persist four-dimensionally in virtue of having numerically distinct temporal parts). I argue that the endurance view of persons is ontologically prior to the perdurance view; on the ground that objects must always be individuated under some substance sortal concept S (the sortal dependency of individuation), and that the concept person entails that objects falling under it are three-dimensional.

In chapter 4-8, I discuss the second problem. I differentiate between Criterianists, who maintain that it is possible to specify a non-circular and informative criterion for personal identity, and Non-Criterianists, who deny that such a specification is possible. In chapter 5-7, I consider in turn Psychological Criterianism, Physical Criterianism and Animalism. I argue that none of these accounts is adequate on the ground that they are either (i) circular, (ii) violate the intrinsicality of identity, or (iii) do not adequately represent what we are essentially. In chapter 8, I discuss Non-Criterianism. I consider in turn Cartesianism, The Subjective view and Psychological Substantialism. Against these accounts I argue that they wrongly assume that ‘person’ refers to mental entities.

In chapter 9, I formulate a biological Non-Criterianistic approach to personal identity; the Revised Animal Attribute View. Person is a basic sortal concept which picks out a biological sort of enduring persons. A person, then, is an animal whose identity as person is primitive in relation to his identity as an animal. I claim that the real essence of a person is determined by the real essence of the kind of animal he is, without thereby denying that persons have a real essence as persons.
Persons–Their Identity and Individuation
Persons—Their Identity and Individuation

Roger Melin

Umeå University 1998
ABSTRACT

This study is about the nature of persons and personal identity. It belongs to a tradition that maintains that in order to understand what it is to be a person we must clarify what personal identity consists in. In this pursuit, I differentiate between the problems (i) *How do persons persist?* and (ii) *What facts, if any, does personal identity consist in?*

Concerning the first question, I argue that persons persist three-dimensionally (*the endurance view*), and not four-dimensionally (*the perduarance view*), on the ground that objects must always fall under some substance sortal concept S (*the sortal dependency of individuation*), and that the concept *person* entails that objects falling under it are three-dimensional.

Concerning the second question, I differentiate between *Criterianists*, who maintain that it is possible to specify a non-circular and informative criterion for personal identity, and *Non-Criterianists*, who deny that such a specification is possible. I argue against Criterianist accounts of personal identity on the ground that they are either (i) circular, (ii) violate *the intrinsicality of identity* or (iii) do not adequately represent what we are essentially. I further criticise three Psychological Non-Criterianist accounts of personal identity on the ground that they wrongly assume that ‘person’ refers to mental entities.

Instead I formulate *the Revised Animal Attribute View* where *person* is understood as a *basic sortal concept* which picks out a *biological sort* of enduring animals. In this, I claim that the *real essence* of a person is determined by the real essence of the kind of animal he is, without thereby denying that persons have a real essence as persons.

Keywords: person; personal identity; criterion of identity; sortal concept; individuation; perdurance view; endurance view; Criterianism; Animalism; Non-Criterianism; animal attribute view; basic sortal; real essence; sort; natural kind.
ACKNOWLEDGEMENTS

I would like to thank my supervisor docent Sten Lindström. He has always taken his time to discuss philosophical matters with me, and has been of invaluable help in structuring my philosophical thoughts. Most ideas in this thesis are, in one way or other, inspired by the discussions we have had over the years. I would also like to thank Professor Martin Edman who was my supervisor in the early stages of the work on this thesis.

I would like to thank Mr Paul Snowdon who supervised me when I was at Exeter College, Oxford during the academic year 96-97. Many ideas which are defended in this thesis are inspired by the discussions we had during that time. I am grateful to the department of Philosophy and Philosophy of Science for enabling me to stay at Exeter College for a year.

I am also very grateful to all participants of the seminars at the department of Philosophy and Philosophy of Science. The comments given to me on earlier drafts of this thesis have been of highest importance for me. Especially, I would like to thank Bertil Strömberg who, besides being my assistant supervisor, also has been of invaluable help in practical as well as philosophical matters. Anders Berglund also needs mentioning. He has always been keen on discussing philosophical questions concerning persons and personal identity. Anders also needs a special mention, since he has helped me with problems concerning the layout of my thesis.

I would also like to thank Jayne Waterworth for helping me make my English more idiomatic.

Most of all, though, I am forever in debt to my family. Without the love, care and patience from my girlfriend Helena and our children Aron, Nils and Isak this work would never have been completed. This thesis is dedicated to all of them, and to all the persons who are no longer with us.
TABLE OF CONTENTS

ABSTRACT
ACKNOWLEDGEMENTS

1. INTRODUCTION .................................................................................. 1
   1.1 Introduction .............................................................................. 1
   1.2 The problem of personal identity over time ......................... 3
   1.3 A methodological remark ..................................................... 10
   1.4 The content and structure of the work ................................. 16

2. IDENTITY AND CRITERION OF IDENTITY .................................. 21
   2.1 Introduction ............................................................................. 21
   2.2 Identity .................................................................................... 22
   2.3 Numerical and qualitative identity ....................................... 26
   2.4 Identity and change ............................................................... 27
   2.5 Cases of fission and fusion .................................................... 32
   2.6 Criterion of identity ............................................................... 34

3. SORTAL INDIVIDUATION AND PERSONS .............................. 41
   3.1 Introduction ............................................................................. 41
   3.2 Sortal individuation, sortal terms and sortal concepts .......... 42
   3.3 Substance sortal terms and phased sortal terms ................. 47
   3.4 Identity, sortal individuation and criteria of identity ........... 49
   3.5 The sortal term ‘person’ ........................................................ 50
   3.6 Ontological priority ............................................................... 52
   3.7 The persistence of persons: the Perdurance view ................. 54
   3.8 Criticism of the perduring understanding of persons .......... 57
   3.9 The persistence of persons: the Endurance view ................. 60
   3.10 The ontologically priority of the Endurance view .......... 64

4. CRITERIANISM AND NON-CRITERIANISM .......................... 67
   4.1 Introduction ............................................................................. 67
   4.2 Criterianism ........................................................................... 68
   4.3 The specifiability thesis ......................................................... 70
   4.4 The complexity thesis .......................................................... 72
   4.5 The derivativeness thesis ....................................................... 74
   4.6 The essentiality thesis .......................................................... 75
   4.7 Different versions of Criterianism .......................................... 76
   4.8 The indeterminacy thesis ....................................................... 77
   4.9 The reductionist thesis .......................................................... 79
   4.10 Non-Criterianism ................................................................. 84

5. PSYCHOLOGICAL CRITERIANISM .......................................... 87
   5.1 Introduction ............................................................................. 87
   5.3 The circularity-objection ....................................................... 93
   5.4 Quasi-memories and personal identity over time .................. 97
   5.5 Criticism of Psychological Criterianism .............................. 102
1. INTRODUCTION

1.1 Introduction

In some sense of the word "know", we all know what it is to be a person that exists over time. I, for instance, know that I am a person and that I went to a restaurant yesterday and had my dinner. I also know that Hansson, who went there with me, is one and the same person as the Hansson I usually have dinner with. Not only that, I also know that Hansson, if questioned about what he did yesterday, would express a similar kind of knowledge of the same fact. He also knows that he is a person existing over time.

This apparent certainty of our identity over time as persons is understandable if we consider the fundamental importance that person and personal identity have for the understanding of ourselves as cognitive creatures. We understand ourselves as able to perform actions and having intentions, as giving and receiving praise and blame, we reflect and have emotions, we have perceptions, feel sensations, we can be punished and rewarded, we acquire and have beliefs and knowledge, etc. Though this list of what we normally think of ourselves as being able to do is far from exhaustive, it shows us that we conceive ourselves to be subjects of experience and agents of actions of which it is appropriate to attribute to ourselves what P. F. Strawson once called 'P-predicates', i.e. predicates by which we attribute, or "imply the possession of consciousness on the part of that to which they are ascribed". A person is an agent of actions and a subject of experience. And since we think of ourselves as agents of action and subjects of experience, the concept of a person is of importance for understanding ourselves.

Conceiving ourselves as the same person over time is of no less importance for our self-understanding than that of conceiving ourselves to be persons. For instance, being the same person seems to be a necessary requirement if the

---

2 Ibid. p. 105
notions of ‘responsibility’, ‘guilt’ and ‘freedom’ are to function as they do. If we did not have the notion of ‘personal identity over time’ it would not be intelligible to claim that a particular person P did the morally wrong action A. This is partly because personal identity seems to be a condition for there being any moral actions at all, since the carrying out of an action spans over a certain period of time. Another reason is that a moral agent, it is supposed, is only responsible for his own actions. Unless we took the person P, who performed the action A at time t, as being the same person as P' who we hold responsible at t' for the action A carried out at t, it would not be appropriate to hold P' responsible for the action at all. Another example has to do with ‘ownership’. We think it to be legally appropriate to claim that if someone owns a certain thing x, then it is only that person who has the right to decide what to do with x in the future.

From the perspective of ourselves as the “knower”, or information-gaining individual, being the same person at different times also seems to be of importance. As the subject of experience the person takes in information from the external world and from his own inner sense, and on that received information the person forms beliefs and knowledge about his internal states of consciousness as well as about the external world. But, as it seems, a person could not acquire such knowledge, or beliefs of internal states and external facts unless he remained the same person over time. This is mostly obvious in cases of self-locating knowledge, but it also appears to ground our understanding of knowledge of facts. Self-locating knowledge is knowledge about the subject’s own particular location at a particular time, such as my knowledge that I went to a restaurant yesterday. In having this knowledge I assume, or believe that I am a person who was at a restaurant yesterday and as such my knowledge about the fact that I was at a restaurant yesterday is grounded in the belief that I am identical with a person who went to a restaurant.

Secondly, it seems that certain kinds of linguistic expressions expressing knowledge of facts presuppose, for their very intelligibility, personal identity over time. Expressions such as, ‘seen before’, ‘once again’ and ‘similar to’ get their significant informational value in virtue of being part of a system of thought in which personal identity lies as a fundamental concept. ‘Seen before’ and ‘once again’ are expressions about re-identification and it seems to be the case that the fact of re-identification is possible only if a person remembers having seen a certain object earlier. That is, re-identification of objects of experience seems to require that it is the same person that experiences one object O

---

Introduction

at two different times \( t \) and \( t' \). ‘Similar to’ also seems to be relying upon the subject being able to remember earlier experiences, and hence, upon a presupposition of personal identity. Suppose a subject has a visual perception of a celestial body. If the subject is to acquire knowledge of that particular celestial body he must draw on earlier experiences to be able to recognise similarities and dissimilarities between this newly observed celestial body and earlier observed bodies. It is only through a comparison of this new visual experience to other, and earlier, experiences that the subject can reach the conclusion that the presently observed celestial body mostly resembles a planet, and that the observed body therefore (probably) is a planet. But, it is difficult to understand how the subject of experience could accomplish the required comparison unless he is the very same person as the one who earlier gained knowledge about similarities and dissimilarities between different kinds of objects.

1.2 The problem of personal identity over time

So far I have considered reasons to suppose that both the notion of ‘a person’ as well as that of ‘personal identity over time’ are of fundamental significance in the way we conceive of and talk about ourselves. Let me now say something about the problem of personal identity.

A first general point is that I will understand the problem of personal identity over time as a way of elucidating what it is to be a person, or what the nature of a person is. The reason for me to take this perspective on the problem of personal identity over time is that the persistence of persons depends, at least partially, upon the essential features of persons.\(^4\) If these essential features does not obtain, then neither can personal persistence. In this sense, the question of personal identity over time and the question of the nature of persons, even though they are distinguishable in thought, are two routes to the same knowledge. Even if it is possible to distinguish between the question of what it is to be a person and the question of what it is for a person to exist over time,\(^5\) I think it is a mistake to think that there is a clear-cut gap between them, since a specification of the conditions for the continuing existence of an object of a kind \( K \) also would, to a large extent, be a specification of the conditions for being an object of the kind \( K \).

---


This last idea is strongly opposed by Baruch Brody. As Brody puts it;

Two dualists about persons might well, for example, disagree about identity-conditions for persons, one holding a body theory and one holding a psychological theory, and two proponents of the bodily theory of personal identity might disagree about the nature of persons, one being a materialist and the other being a dualist.

According to Brody, 'identity' is purely general in its application and as such it cannot differentiate between different kinds of objects. The reason for this is that the establishing of natures, or essences of objects, requires a specification of what distinguishing features objects of a certain sort has. But, since 'identity' is purely general in its application it is not sensitive to such distinguishing features and cannot determine any particular features of objects of different sorts. Brody calls questions concerning identity "disputes about identity-questions" and questions of the nature of objects he calls "disputes about the nature of the objects in question".

Though I certainly agree with Brody in the claim that identity is univocal, and as such applicable in the same way to objects of all kinds, I do think his denial of the possibility of illuminating what a person is through answering questions of what personal identity consists in rests upon a misunderstanding of questions of identity. Establishing identity-conditions for persons is not, as Brody seems to think, specification of a certain relation of identity which is said to hold only for persons. Specifying identity-conditions for persons is to establish the persistence-conditions for persons. Determining a question of personal identity is to determine what to count as the continuing existence of a person and what to count as the cessation of a person. But, determining these conditions would also give us an answer to the question what a person is. Since a very fundamental idea associated with 'person' is that of persons existing over time and that the specified conditions would consist of just those features of a person that delineate a particular person from other kinds of things and from other persons existing at the same time. Thus, by getting a clear and perspicuous understanding of what it is, or means, to persist as a person, we would at the same time receive a clear and perspicuous understanding of what it is, or means to be a person.

6 Brody, Identity and Essence, pp. 3-6.
7 Ibid. p. 6.
8 Ibid. p. 6.
9 Ibid. p. 6.
A second general point is that the problem of personal identity can be differentiated into two different, but inter-related problems. I started this chapter with the claim that we know that we exist over time as persons, although I qualified the claim by talking about 'knowledge' in a certain sense. This "knowledge" of ours is grounded on a psychological fact, i.e. that persons have certain psychological capacities enabling them to view themselves as persisting objects. These psychological capacities of persons enable them to have first-person experiences of themselves as being one and the same person over time. Some philosophers have tried to define 'person' in terms of this psychological fact.

John Locke, for instance, defined a person as that object which is "a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places". Let us say that this definition of 'person' tries to specify the first-person perspective condition for persons. According to this line of reasoning, it is a necessary condition for an object \( x \) to be a person that \( x \) can conceive of himself as existing over time. I do not think that the first-person experience condition as it is stated by Locke is correct, but, for the time being let us ignore the question of the correctness of the condition. What is of interest, though, is the fact that what is contained in the first-person perspective, or the informational content given in the first-person perspective illuminates the two problems of personal identity. Regardless of whether all persons necessarily satisfy the first-person perspective condition or not, it is, at least, in virtue of having the first-person perspective that a typical person (such as you and I) view himself or herself as one and the same person existing over time. That is, this psychological capacity enables the typical person to be presented to himself as an object with a certain definite history and a certain, though unknown, future. We can express this thought by the claim that the first-person perspective held by persons presents personal persistence as a fact of the world. By this I mean, firstly that the first-person perspective presents the person as a fundamental unit of predication who exists with different proper-

---

11 This concerns not only persons, but all kinds of objects. For instance, it is a fact that I know that ordinary material objects like books, trees, animals and even snowflakes exist over time. To have an existence over time is nothing peculiar for persons, enduring existence extends to all possible kinds of observable entities.


13 There are many infants, and senile individuals who do not satisfy the first-person perspective condition, but who we still think it appropriate to call persons. But, this does not mean that the first-person perspective condition cannot be a part of a necessary condition, what this show us is only that the first-person experience condition in itself is not a necessary condition for an object to be a person.
ties at different times, i.e. as a *substance*, and secondly, that the presentation of sameness over time is experienced by a person as something immediately given or *primitive*.

But now, the two problems of personal identity emerge. We know that the first-person perspective presents the person as persisting as identical over time, as well as something primitive. But we do not know whether the first-person perspective adequately represents what personal persistence consists in. That is, we do not know the answers to the two semantical-*cum*-ontological questions:

(i) How do persons persist over time?

(ii) In what facts, if any, does personal identity consist?

The first problem concerns whether we should understand ‘personal identity over time’ as implying numerical identity, i.e. as implying the existence of an individual who is wholly present at every time where that individual exists, or if it only implies a weaker relation between qualitatively similar but numerically distinct objects. The second question concerns whether it is possible to give a non-circular specification of the relation of personal identity conceived in the first-person perspective in terms of a continuity-relation other than identity.

Concerning the question of how persons persist we can either claim that persons *perdure* or *endure*. If we claim that persons perdure, then we deny that the first-person perspective adequately presents our existence over time. Persons are, according to this view, four-dimensional objects that persist in virtue of a special relation holding between their temporal parts. One reason for taking this view is certain difficulties associated with the interpretation of ordinary talk about objects remaining the same through change. We cannot, it is assumed, consistently mean *numerical identity* when attributing identity over time ordinary objects or persons, since persistence over time involves change which is thought to be incompatible with numerical identity. What is meant by persisting, according to the understanding of persons as perduring, is always either the succession of a sequence of interrelated experiences, or the succession of a sequence of person-stages. We are, then, neither from an ontological nor a

---

Introduction

If we, on the other hand, claim that persons persist by being enduring objects, then we affirm the idea of numerical identity over time given in the first-person perspective. Persons are, according to this view, three-dimensional continuants who persist through change in virtue of having different properties at different times. Statements about personal identity over time are to be interpreted literally as involving only one object existing at different times. This means that only one object is involved in a case of personal identity over time, and that that object is present at several different times.

The second problem concerns what our identity over time consists in. Broadly speaking, we can discern two possible positions; we can advocate either Criterianism or Non-Criterianism. A theory is Criterianist with regard to a predicate ‘F’ if what ‘F’ refers to is understood as being semantic-cum-ontologically complex and derivative in relation to some other facts T₁,..., Tₙ. Non-Criterianism with regard to a predicate ‘F’, on the other hand, maintains the view that what ‘F’ refers to is semantic-cum-ontologically primitive in relation to all other facts T₁,..., Tₙ. The central claims of Criterianism with regard to the predicate ‘personal identity’ are (i) the fact of personal identity over time consists in the holding of certain other facts, (ii) it is possible to specify informative necessary and sufficient conditions for personal identity in terms of these more particular facts, and (iii) these particular facts that are specified should pick out what essentially is involved in a case of personal identity over time. Non-Criterian-

---

17 The perduring view which I am discussing, then, is a semantic-cum-ontological position. I am not debating the fact that we can represent identity over time of persons as a succession of interrelated stages once the ‘numerical identity of persons’ has been taken as the primitive notion. This means that I will not discuss philosophers who accept the semantic-cum-ontological primitive nature of strict personal identity and who think we can describe this numerical sameness of a person as a relation between states of the person. See for instance Eli Hirsch, The Concept of Identity, (New York: Oxford University Press, 1982).

18 In classifying the theories of personal identity into Criterianism and Non-Criterianism I am using a taxonomy that differs in certain details from the traditional taxonomies associated with theories of personal identity: That of the distinction between the "Simple View" and the "Complex View" and the distinction between "Reductionism" and "Non-Reductionism" (See Derek Parfit, ‘Personal Identity and Rationality’ from Synthese, vol. 53, No. 2, 1982, and Reasons and Persons, ). The reasons for me to change the taxonomic system is that the proposed distinction, though capturing certain fundamental differences between the two opposing traditions of viewing personal identity over time, are to a certain extent unclear and that they cannot capture what I understand to be the essential question; whether it is possible to specify a criterion for personal identity or not.

ism with regard to the predicate ‘personal identity’ maintains that personal identity is a primitive unspecifiable further fact, since a person’s existence over time must be understood as something other than the existence of all other facts of the world.\textsuperscript{20}

The reason for inquiring into the problems of what it is to be a person who persists over time is that it is a commonly known fact that it is possible to describe cases in which (i) we are uncertain whether a particular object is to be counted as a person or not, and (ii) we do not seem to know whether we have a case of personal identity or not. Such described cases show us that our intuitions concerning what a person is and what personal identity consists in are not as clear and perspicuous as we normally take them to be. Sometimes our intuitions concerning such cases lead in different directions, and leave us bewildered. But such “puzzle-cases” are disturbing. Not knowing what to think of a case creates what Bernard Williams calls a “conceptual shadow”.\textsuperscript{21} However, if a conceptual shadow always accompanies talk and thought about persons and personal identity, and such talk and thought are of fundamental importance for understanding ourselves as acting and knowing individuals of the world, then it seems that we would never be able to reach any self-understanding at all. This means, then, that getting knowledge of what it is to be a person, and what personal identity consists in is of primary interest if we want to know what kind of entities we are, and a theory that would increase our self-understanding by giving a clear and perspicuous account of those notions is worth pursuing.

Take, for instance, the case of split-brain patients.\textsuperscript{22} Humans have certain nerve fibres, the \textit{Corpus Callosum} and some smaller pathways which integrate information of the left hemisphere with information of the right hemisphere.\textsuperscript{23} The left hemisphere is mainly associated with processes of the right side of the body and the right side of the visual field, while the right hemisphere is mainly associated with processes in the left side of the body and the left side of the visual field.


\textsuperscript{22} Thomas Nagel, \textit{Brain Bisection and the Unity of Consciousness}’, from \textit{Mortal Questions}, Canto edition, (New York: Cambridge University Press, 1991), pp. 147-164. Other examples are the possibility of intelligent parrots, teletransporters and "mind-extracting devices".

\textsuperscript{23} \textit{Ibid.} pp. 147-154. The transmission of vision is actually ipselateral, since it is the left side of the retina which takes in information from the right side of the visual field, and the right side of the retina which processes information from the left side of the visual field.
Besides processing information of the right side of the body (including the visual field) the left hemisphere usually controls the production of speech. It is possible to treat severe epilepsy through cutting off the main integrational function between the hemispheres of the brain. These patients are split-brain patients. A consequence of the lack of integration between the two hemispheres is that a split-brain patient acts in certain experimental situations as if there are two different streams of consciousness going on at the same time in one body. It was, for instance, reported that

[W]hat is flashed to the left half field or felt by the left hand cannot be reported, though if the word 'hat' is flashed on the left, the left hand will retrieve a hat from a group of concealed objects if the person is told to pick out what he has seen. At the same time he will insist verbally that he saw nothing (...) 25

and if a split-brain patient has a hidden object in his left hand he will show the correctness or failure of a guess of what kind of object it is by his facial expression.26

But, now, what are we to say about a split-brain case, and particularly, how many persons are involved in such a case? Is a split-brain patient one person, or two?

Another example is the problem of euthanasia. Most of us have the intuition that suffering is something bad. We also feel a certain aversion towards killing another person. If we then consider a case where we have an individual that suffers badly, and the only way to end the suffering is to end that individual's life, would it then be morally wrong of us to help that individual to end the suffering by helping the individual to die? If the object is not a person, such as a dog or a cat, then most of us would think we have an obligation to help that individual to die. But if the individual is a person, what then should we do?27

Though questions such as this await an answer, I do think it cannot be given until one has a firm understanding and determination of what it is to be a person.

---

24 The exception to this contralateral transmission is tactile information of the head and the neck, auditory information from the ears and smells. Information from the head and the neck, and auditory information are connected to both hemispheres at once, while smells passes from the left nostril to the left hemisphere and the right nostril to the right hemisphere.


27 It should be noted that I am not involving myself in a direct discussion of ethical matters in this work. My aim is to present a theory of personhood. This account rests entirely upon understanding 'person' as an ontological category of objects on a par with other categories of objects such as tigers, trees, tables and cars.
1.3 A methodological remark

The use of thought experiments is extensive in the literature on personal identity, and as far as I know, no other area of philosophy makes such heavy use of thought experiments. There are thought experiments involving teleportation, brain-transplants, bionic robots, and the possibility of disembodied existence and reincarnation. But, even though thought experiments are of frequent use in the literature, not everybody is happy about that fact. Some philosophers have raised doubts whether thought experiments are reliable sources of information upon which we can ground any truths about persons and personal identity. Let me first give a brief description of what I take to be essential features of thought experiments.28

Roughly speaking, a thought experiment can be expressed as a hypothetical conditional, where the antecedent of the conditional is a description of a certain imagined, or possible situation and the consequent of the conditional describes how we would think, talk, perceive or act given that the described imagined situation were actual. Since thought experiments are imagined experiments conducted in the mind of an experimenter a thought experiment need not, contrary to an ordinary experiment, be bound by empirical constraints.29 The important fact of the thought experiment is that the imagined situation is supposed to be a situation occurring in a world resembling our own, with the exception of the specified event of the thought experiment. The imagined situation together with the supposition of its occurrence in a world resembling ours in every relevant aspect is a complete determination of all truth-value conditions and it would therefore make the thought experiment to be a description of a possible world. Thus, a thought experiment consists of a description of a possible world and the implications of how we would think, talk, perceive and act given that the possi-

28 For a thorough examination of thought experiments, see Sören Häggqvist, Thought Experiments in Philosophy, Published by Stockholm University, (Stockholm: Almquist & Wiksell International, 1996).

29 Contrary to ordinary scientific experiments in which a certain situation is actually realised in a laboratory, thought experiments are, as the name indicates, experiments only imagined in the mind of an experimenter. This imaginative part of thought experiments is principally connected to thought experiments. It is not enough for an experiment to be a thought experiment if a certain individual only imagines the conducting of an ordinary scientific experiment. Not only should the experiment be an imagined experiment, it should also be an experiment consisting of an imagined situation. But, even though thought experiments are conducted in the mind of an experimenter, thought experiments are real experiments, since they are conducted as tests of theories, theses or concepts.
ble world would be the actual world. It is in these implications of the imagined situation, or the possible world that the thought experiment, or "the method of cases" as Mark Johnston calls it, has its informational value to our understanding of our actual concepts and theories. For instance, if we in the actual world have a certain concept $C$, and we want to know what is conceptually entailed in the concept $C$, or what it is to be an object $c$, we can illuminate this by describing an imagined situation containing a phenomenon relevant to the use of $C$, but which is different from the actual world in some respect, and then "see" whether we would be prepared to apply the concept $C$ to that phenomenon in the imagined situation. By drawing the boundaries of such imagined situations and phenomena to which $C$ is applicable, we would learn about what it essentially takes to be a $C$.

The criticism of thought experiments can be divided into mainly four different strands of argumentation:

(i) imaginability is no guide to metaphysical possibility,
(ii) the described situation is insufficiently specified,
(iii) the described situation is impossible or abnormal,
(iv) intuitions are unreliable.

The first criticism consists in claiming that the fact that we have been able to imagine a certain situation does not establish that that situation is a metaphysical possibility. It is one thing, it is claimed, for something to be a logical possibility/impossibility and a completely different thing for something to be a metaphysical possibility/impossibility. Intuitively speaking, using 'logical possibility' commits one to the existence of the largest set of possible worlds. This is so, because logical possibility only concerns what we meaningfully can conceive, describe or imagine without invoking ourselves in a contradiction. As long as an imagined situation is in accordance with the laws of logic, i.e. it is not self-contradictory, it is regarded as a logical possibility. Metaphysical possibility/impossibility, on the other hand, is a much more restrictive use of 'possibility'. Metaphysical possibility depends upon the real natures of the objects in the actual world. What is metaphysically possible is, in this sense,

sensitive to empirical investigations in the actual world. Hence, the set of metaphysically possible worlds is a subset of the set of logically possible worlds.\textsuperscript{31}

For instance, given that water is composed of $\text{H}_2\text{O}$, it is metaphysically impossible that water could consist of another structure than $\text{H}_2\text{O}$. To see the plausibility of this claim just imagine what we would say if we, in the actual world, found a liquid morphologically resembling water, but which we found to consist of some radically different chemical structure than $\text{H}_2\text{O}$. In this case we would insist that the newly found liquid is something other than water on the ground that its chemical composition is different from water. The same would also be the case when we consider possible worlds. A liquid morphologically resembling water existing in a possible world would simply not be water if it is chemically distinct from $\text{H}_2\text{O}$. But, of course, it is nevertheless conceivable, i.e. logically possible, that water might have had another structure than $\text{H}_2\text{O}$. We can imagine that water does not, as it actually does, consist of $\text{H}_2\text{O}$, i.e. it is not a logical contradiction to suppose that water actually had another chemical composition than the one it actually has. In this sense, then, “Conceivability is no proof of possibility”.\textsuperscript{32}

But, since what is metaphysically possible need not coincide with what is imaginable, the fact that we can imagine a certain situation cannot be taken as establishing that that situation is a metaphysical possibility. Hence, use of thought experiments is not a reliable source for determining what personal identity consists in.\textsuperscript{33}

The second strand of criticism consists in claiming that philosophical thought experiments are unreliable guides to conclusions, since they are underdescribed.\textsuperscript{34} For instance, if a situation is described as $S$, we might infer the conclusion $C$. But, the same situation $S$ could be described with further details of

\begin{footnotesize}
\begin{enumerate}
\item This does not, as it seems, only concern the relation of conceivability and metaphysical possibility. It also seems to be true of possibility, \textit{simpliciter}. One such case is Escher-type pictures. In such pictures it seems as if we are conceiving of a certain situation which is, as closer investigation reveals, impossible. What we imagine cannot be taken to establish a genuine possibility, since what we imagine need not be clearly thought through. Either it could be the case that the thought experiment is underspecified and if we were to elaborate it we should see that it contains logical contradictions, or it might be the case that we have a perfectly complete description which also is perfectly coherent, but which simply cannot be metaphysically realised.
\end{enumerate}
\end{footnotesize}
the background, B, and if S is described together with B it might be the case that we infer the conclusion not-C. Hence, the outcomes or conclusions of thought experiments are heavily dependent of how detailed the description of the thought experiment is.

An example of an underdescribed though experiment is the thought experiment in which persons are supposed to be able to split like amoebas. As Kathleen Wilkes has pointed out, if we are able to infer a conclusion from evaluating this thought experiment we need to have further information concerning such things as how often a person splits, whether the splitting of a person is something we can predict, whether we can prevent splitting, and what the society of this possible world is supposed to look like?

The third strand of criticism against the use of thought experiments is continuous with the criticism of underdescription. Suppose that we are able to give a detailed characterisation of the background information of the thought experiment. Such a description can describe a situation which is very unlike anything that occurs in the actual world. But, now, since our terms and concepts are not created to be applicable to abnormal cases, it seems that a thought experiment, and especially thought experiments with abnormal situations, is without significance for what we mean by our terms. As Wittgenstein said

it is only in normal cases that the use of a word is clearly prescribed; we know, are in no doubt, what to say in this or that case. The more abnormal the case, the more doubtful it becomes what we are to say. And if things were quite different from what they actually are (...) this would make our normal language-games lose their point.35

W. V. Quine expressed the same idea concerning the use of science fiction in the literature on personal identity:

To seek what is “logically required” for sameness of person under unprecedented circumstances is to suggest that words have some logical force beyond what our past needs have invested them with.36

The fourth strand of criticism against using thought experiments clams that thought experiments are unreliable because the conclusions of thought experiments are based upon unreliable intuitions. The reasons that intuitions are supposed to be unreliable are (i) that they are based upon subjectivity in that they are sensitive to the background assumptions and background knowledge of the


persons having the intuitions, and (ii) that they are heavily dependent upon the linguistic description of the thought experiment. This last idea has been especially emphasised by Mark Johnston.37

Consider the following two thought experiments originally designed by Bernard Williams:38

(i) Imagine that there is a process in which all information contained in a person’s brain is extracted and stored in a mind information device. Consider further that there are two different persons, A and B, who have their information extracted in this sense, and that we later induce the A-information in the B-body and the B-information in the A-body. Let us call the person with A’s previous body the A-body person and the person with B’s body the B-body person. Suppose now that before the process of extraction gets started, A and B are told that one of the resultant persons will receive a reward, say one million dollars, while the other will be tortured and that A and B are to egoistically choose which treatment should be given to which resultant person. Let us suppose that A chooses the reward to be given to the B-body-person and the unpleasant torture to the A-body-person and B chooses the conversely. Suppose the experimenter gives the torture to the B-body-person and rewards the A-body-person. Will not, then, the B-body-person complain that this outcome was not what he chose and that his choice was that the torture should not happen to him. And, will not the A-body-person feel comfort in the fact that the experimenter’s choice followed his wishes. But, citing Williams:

These facts make a strong case for saying that the experimenter has brought it about that B did in the outcome get what he wanted and A did not. It is therefore a strong case for saying that the B-body-person really is A, and the A-body-person really is B; and therefore for saying that the process of the experiment really is that of changing bodies39

and that

the results suggest that the only rational thing to do, confronted with such an experiment, would be to identify oneself with one’s memories, and so forth.40

(ii) Imagine that a person A is a prisoner of a mad scientist. The scientist tells A that he will conduct a very painful experiment on A tomorrow. As a result of

38 Williams, ‘The Self and the Future’, pp. 46-52
39 Ibid. p. 49
40 Ibid. p. 51
that A, naturally, fears the sunrise of the next day. But the scientist further says that A need not worry about the experiment, since the scientist will extract the information contained in A's brain before the experiment begins, and that he will induce the brain with completely new information during the painful experiment. Only after the painful experiment has been completed will he re-induce the original information of A's brain to A. But, of course, this cannot ease the anxiety A feels concerning the experiment; it more probably exaggerates the anxiety. Not only is he to be enduring a painful experiment, his brain is also "to be fiddled with in a fairly drastic way so as to produce radical psychological discontinuity". So, from A's egoistic point of view, it seems to be rational that he tries to prevent the conducting of the experiment, since it is he who will undergo pain. But, this means that psychological continuity between a person A and a future pain F is not required for the future pain to be A's pain. It seems more likely that it is body continuity which is required for such a relation to obtain.

Johnston's point is now that it seems to be the case that these two thought experiments really are two different descriptions of one and the same case, and that "the two intuitions taken together threaten the method of cases". Since the only ground for having different intuitions concerning these two thought experiments can be how the "fact" of the thought experiment is presented to us, this means that our intuitions are sensitive to the description, and that the conclusions intuitively drawn from thought experiments are biased, or prejudiced in favour of how the thought experiment is presented to us. But, if intuitions are sensitive to the description of the thought experiment, how can we ever rely upon the inferences drawn from thought experiments?

I do think that all of these criticism against the use of thought experiments are, to a certain extent, correct. But, this does not imply that we cannot use thought experiments in determining what essentially is involved in cases of personal identity. As Derek Parfit once wrote, the criticism against the use of thought experiments might be justified if, when considering such imagined cases, we had no reactions. But these cases arouse in most of us strong beliefs. And these are beliefs, not about words, but about ourselves. By considering these cases, we discover what we believe to be involved in our own continued existence, or what it is that makes us now and ourselves next year the same people. Though our beliefs are revealed

---

41 Johnston, 'Human Beings', p. 360. We can imagine the mad scientist in the second thought experiment using a mind-extractor-device like the one used in the first thought experiment.

42 The only difference between the two cases are that B's experiences are left out of the description in the second case.

43 Johnston, 'Human Beings', pp. 360-361
most clearly when we consider imaginary cases, these beliefs also cover actual cases, and our own lives.\textsuperscript{44}

In this sense, then, thought experiments can, and do, give us knowledge about what it is like to persist as a person. But, it does this in a fallible way. What is needed, then, is not a dogmatic rejection of thought experiments as such, but a natural and critical scepticism towards particular thought experiments. We cannot, and should not think of thought experiments as absolutely guaranteeing the truth of the conclusions. A thought experiment taken in isolation is a rather poor argument for a claim. But, this is as it should be. It is only when a thought experiment is taken in relation to a whole collective body of other arguments that it receives its significance. The received knowledge from a thought experiment is, then, fallible. We cannot be absolutely certain that the inference drawn from a thought experiment is true. But, what we can do is to establish that an inference drawn from a thought experiment is plausible. If we have a consistent system of arguments claiming the thesis V, where each argument is plausible relative to its competitors, and a thought experiment confirms V, then we can claim that the conclusion of the thought experiment is plausible and that the thesis V is plausible relative to other possible “theories”.

1.4 The content and structure of the work

In this book I am going to argue for and present a particular account of persons and personal identity. It is an account which has as its objective to determine what it is to be a person from the perspective of how we basically have used and do use the terms ‘person’ and ‘personal identity’. I will defend the thesis that the concept \textit{person} is a basic sortal concept which picks out a natural biological sort of enduring persons. A person is, according to this view, an object of the world which remains the same over time in virtue of being numerically the same at every moment of its existence and which belongs to the sort \textit{Person} in virtue of having certain essential properties.

In chapter 2, I will elucidate the two notions ‘identity’ and ‘criterion of identity’. I will claim that ‘identity’ stands for, or ought to stand for, the conceptually primitive \textit{numerical identity}. I then consider, and refute two objections towards my interpretation of ‘identity’: (i) the claim that numerical identity over time is incoherent, and (ii) that cases of fission and fusion show us that we cannot always mean \textit{numerical identity} by ‘identity’.

\textsuperscript{44} Parfit, \textit{Reasons and Persons}, p. 200.
I will further claim that 'criterion of identity' should be understood as a semantic-cum-metaphysical determination of what it means, or takes to be an object of a certain sort, or kind. I end the second chapter with a discussion of the thesis of Relative Identity. I claim that the notion of 'relative identity' presupposes a prior understanding of absolute identity and that purported counterexamples against the concept of absolute identity can be reformulated in ways that make them consistent with the absolute understanding of identity.

In chapter 3, I will defend a conceptual and ontological version of the thesis of the sortal dependency of individuation. This thesis maintains that: (i) there cannot be any individuation of objects without the objects falling under some sortal term 'S'; (ii) this sortal term 'S' provides a principle of individuation for objects of the sort S, i.e., there is a determinate answer to the question for any two objects a, b, of the sort S whether a is the same S as b; and (iii) if a, b are objects of the sort S, then a is identical to b if and only if a is the same S as b. Thus, sortal terms determine how many objects of a certain sort S there is in a region of space during a certain period of time.

I will also claim that 'person' is a sortal term and that we have prima facie evidence for thinking it to be a substance sortal term, i.e., a term that is applicable to a person throughout its entire existence or not applicable at all. It is a part of our concept of a person that persons persist through time in accordance with their experiences of their own and other’s’ identity. Persons are therefore enduring objects, not perduring objects like events and processes. This is "established" by showing that there are certain problems with the perdurance view, and showing that this view is conceptually and ontologically dependent upon the endurance view of ourselves.

In chapter 4, I discuss the general distinction of Criterianism and Non-Criterianism. This chapter is purely descriptive. The general point of Criterianism is that Criterianist accounts of personal identity accept that it is possible to specify, in a non-circular way, necessary and sufficient conditions for personal identity over time, since the fact of personal identity depends upon some more "basic" facts. A Criterianist might also accept that some questions of personal identity might lack a determinate answer and/or that it is possible to reduce all talk about persons to talk about some more basic facts.

Non-Criterianism, on the other hand, maintains that personal identity cannot be non-circularly specified in terms of necessary and sufficient conditions, and that personal identity is primitive in relation to the empirical continuities that might underlie it.
Chapter 5, contains a presentation and criticism of Psychological Criterianism. The general idea of Psychological Criterianism is that psychological continuity of some form of other is the decisive condition for personal identity. I distinguish between two different versions of Psychological Criterianism: Locke's original version and the Psychological criterion. I will argue, following an objection formulated by Joseph Butler and Thomas Reid, that both of these accounts of personal identity are inadequate on the ground that neither one of them can be non-circularly specified.

In chapter 6, I discuss theories which claim that personal identity consists in spatio-temporal continuity of a physical object. I divide these accounts into the Brain criterion, the Physical Realiser criterion and the Body criterion. I will claim that all Physical Criterianism accounts of personal identity are inadequate, since they entail the possibility that a person's identity over time can depend upon the simultaneous existence of another person. A consequence of my criticism is that Physical Criterianism cannot be understood as specifying a sufficient condition for personal identity, even though it could be the case that some version of Physical Criterianism specifies a necessary condition.

In chapter 7, I discuss the view that the identity over time for persons like you and I consists in sameness of animal, viz., Animalism. Two different versions of Animalism will be described: one which understands the terms 'person' and 'human being' as equivalent, and one which understands 'person' as a phased sortal term applicable to objects falling under, amongst others, the substance sortal term 'human being'. I will claim that the latter form of Animalism is the most plausible one on the ground that we cannot, a priori, rule out the possibility of persons other than human beings. I will further claim that neither form of Animalism could be correct on the ground that (i) any form of Animalism already employs a pre-understanding of personal identity over time, and (ii) that the second, and most plausible version, relies upon a misunderstanding of what we are essentially, or what substance sortal it is that picks out what we are essentially.

In chapter 8, I will describe and criticise Psychological Non-Criterianist views, and particularly, I will describe and criticise Cartesianism, the Subjective View and Psychological Substantialism. What these views have in common is that they, from a mainly first-person subjective perspective of personal identity and what it is to be a person, conclude that persons primarily are objects of a psychological character. I will claim that neither one of these positions is adequate. The first two because persons are, by necessity, embodied objects, since the ability of an object to have properties normally associated with personhood
relies upon the fact that the object also has certain corporeal characteristics. I further argue that both of these accounts rest upon three Cartesian mistakes.

Against Psychological Substantialism I claim that it is not reasonable to think persons to be psychological substances, and hence *person* to be a psychological substantial sort. The reasons for this are that individuals like you and I need not have psychological capacities at all times that we exist, and (ii) that we have reasons to suppose that animals can have the same higher cognitive capacities as persons. Since the higher cognitive capacities of animals should be biologically explained it seems reasonable to suppose also that the higher cognitive capacities of persons should be biologically explained.

Chapter 9 will consist of conclusions and inferences based upon the earlier chapters. In chapter 9, I claim that one must understand *person* as a basic sortal concept picking out a substantial sort of object in the world. I discern this claim into five different parts; (i) that *person* is a primitive concept by which objects are individuated, (ii) that personal identity is primitive in relation to the empirical conditions that underlie it, (iii) that persons are objects that endure over time in the sense that individual persons persist in virtue of being wholly present at each time of their existence, (iv) that 'person' refers to a substantial sort *Person* of objects in the world, and (v) that 'person' is a substance sortal term in that it picks out an essential feature of individual persons. I further claim that the so called ‘animal attribute view’ is correct in claiming that a necessary condition for being a person is to belong to an animal kind. I claim, contrary to the animal attribute view, that persons have a real essence *qua* persons and that *person* is a biological cross-classificatory concept. This view, which I call ‘the Revised Animal Attribute View’, specifies a necessary condition for being a person:

If *a* is a person, then *a* is an animal which falls under the extension of a kind with a distinctive real essence which, if specifiable, would give an account of the ability of a fully evolved individual of the kind to have higher cognitive and social capacities, such as the ability to perceive, remember, imagine, caring for individual of its own or like species, and who is able to conceive itself as existing over time.

In chapter 9 I also investigate the question of whether personal identity is indeterminate. Though I accept that a certain form of quasi-indeterminacy of personal identity is possible due to the fact that persons are subject to evolution, I deny genuine indeterminacy to be possible on the ground that (i) the meaningfulness of attributing personal identity to a pair of objects relies upon a prior individuation of the objects as persons, and (ii) there exists a formal argument
for the conclusion that the relation of indeterminate identity should be understood as something different from the relation of identity.
2. IDENTIFY AND CRITERION OF IDENTITY

2.1 Introduction

In this chapter I will study the two notions 'identity' and 'a criterion of identity'. The reason for doing this is that a general account of those notions underlies, as a conceptual presupposition, a thorough understanding of personal persistence. For a person P to persist from time $t_1$ to time $t_2$, where $t_1$ is distinct from $t_2$ the following two things seem to be required: (i) that there exists a person $P^*$ at the time $t_1$ and there exists a person $P'$ at the time $t_2$, and (ii) that the person existing at $t_2$ is the same person as the one existing at $t_1$.

I will claim that 'identity' always stands for, or always ought to stand for, numerical identity, and that numerical identity is conceptually primitive in the sense that it cannot be defined in terms of any more fundamental concepts. Though numerical identity is a primitive concept, I still think it is possible to establish some necessary conditions for identity. After the establishing of necessary conditions for identity, I investigate the difference between numerical identity and qualitative identity. I will claim that 'qualitative identity' is not a kind of identity at all, but that it stands for the relation of (exact) similarity. I then consider and refute two objections towards my interpretation of 'identity' as always meaning numerical identity: (i) the claim that numerical identity over time is incoherent, and (ii) that cases of fission and fusion show us that we cannot always mean numerical identity by 'identity'. Then I consider what we mean by 'a criterion of identity', and my conclusion is that 'a criterion of identity' should be understood as a semantic-cum-metaphysical determination of what it means, or takes to be an object of a certain sort, or kind. In this sense, a criterion of identity for objects of a sort $S$ is an attempt to give a specification of what it is to be an object of the sort $S$ through specifying the essential conditions for being objects belonging to $S$. Finally, I will discuss the thesis of Relative Identity. I claim that relative identity is grounded on a prior understanding of absolute identity and that purported counter-examples against the idea of absolute identity can be reformulated in ways which makes them consistent with the absolute understanding of identity.
2.2 Identity

The first important feature of ‘identity’ is that it is universally applicable. This means that ‘identity’ is applicable to every existing object, since, by necessity, every object is identical to itself, and to no other object than itself.

That identity is the universal relation which every object necessarily bears to itself cannot, though it is trivially true, be understood as a definition of identity. A minimal requirement of a definition is that it should be non-circular. The *definiendum* should not be a part of, nor a conceptual presupposition for the understanding of the *definiens*. Now, since the notions ‘other’ and ‘other than’ cannot mean anything other than *not being the same*, or *non-identical with*,\(^{21}\) this minimal condition of non-circularity of a definition is violated in the above characterisation of ‘identity’. This fact of a conceptual presupposition of *identity* is a general feature of our thinking and implies that *identity* must be understood as a “conceptual primitive”\(^{22}\), i.e. a concept such that if one did not understand what it is, then one could not possibly have it explained to oneself.\(^{23}\) But, claiming that *identity* is conceptually primitive does not imply that we cannot say anything at all about it. It only means that ‘identity’ cannot be defined in terms of some necessary and sufficient conditions which are conceptually more fundamental than *identity*. Particularly, I think it is possible to put forward a general theory of identity in which necessary conditions of the relation of identity can be established. ‘Identity’ must, of necessity, satisfy the following conditions of standard logic:

1. \(\forall x(x=x)\)  
   *Reflexivity*

2. \(\forall x\forall y((x=y) \supset (y=x))\)  
   *Symmetry*

3. \(\forall x\forall y\forall z((x=y) \land (y=z) \supset (x=z))\),  
   *Transitivity*

4. \(\forall x\forall y((x=y) \supset \forall F (Fx \equiv Fy))\).  
   *Indiscernibility of Identicals, or Leibniz Law*

---


\(^{22}\) *Ibid.* p. 23

The fact that identity is a relation with the formal character of reflexivity, symmetry and transitivity make identity an *equivalence-relation*. The forth condition of identity, i.e., the *Indiscernibility of Identicals* makes it plain that identity is a special kind of equivalence-relation. This Leibnizian principle shows us that identity necessarily implies complete congruence of properties.²⁴ That is, if we for instance say that a person P* is identical with a person P, then the person P* must have the same properties as the person P. Many equivalence-relations do not have congruence as a consequence. For instance, 'being of the same length as' does not.

But, now, someone might wonder, why accept that 'identity' is conceptually primitive. Why only claim that complete congruence of properties is a consequence of identity, why not also claim that complete congruence of properties entails identity. The idea here is that by first accepting the trivial condition:

4. \( \forall x \forall y ((x=y) \supset \forall F(Fx \equiv Fy)) \),

and then, accepting the converse condition:

5. \( \forall x \forall y (\forall F (Fx \equiv Fy) \supset x=y) \)  

we would get the biconditional of (4) and (5):

6. \( \forall x \forall y (x=y \equiv \forall F(Fx \equiv Fy)) \)

which would be a definition of 'identity'.

The idea that we can give a definition of identity has been supported by an argument aiming at showing that the Identity of Indiscernibles is a sufficient condition for identity. Baruch Brody, for instance, has claimed that the following "trivial" argument is enough to establish that it is sufficient for \( a \) and \( b \) to be identical if \( a \) has the same properties as \( b \).²⁵ Brody claims:

(i) Suppose that \( a \) and \( b \) have all of their properties in common
(ii) \( a \) certainly has the property of-being-identical-with-\( a \)
(iii) so, by supposition, does \( b \)
(iv) then \( b = a \).²⁶

Though, this argument does look convincing, I do not think it establishes a sufficient condition of identity that can lie as a foundation of a definition of identity. But, what is it, more exactly, that is wrong with this supposed defini-

---

tion of identity? Several objections have been raised against the definition. For instance, it might be argued that it is too trivial to be a definition, since we have allowed the property of-being-identical-with-itself in the class of properties. Doing this, it will of course be the case that one property, for instance that of a-being-identical-to-itself, if we talk about the object a, will not be included in the class of properties applicable to anything other than a. But, this objection is, I think, pointless. Why should triviality be an argument against a definition? Especially when the supposed definition is of a very fundamental concept within our conceptual schema.

A second possible objection is that there simply is no such property as being-identical-with-a. A property, it is supposed, must be able of plural instantiation, i.e. that it should be possible that more than one object have the property. But, since the property of-being-identical-with-a only can be instantiated by one object, i.e. by a, it is no property at all. I do not think that this objection is sound on the ground that the idea of a property only having one instance involves any contradiction. For instance, we can, and do, talk about the property of-being-the-first-man-on-the-moon. There is only one object that can possess that property, and it seems pretty clear to me that this is something that Armstrong actually possesses. So, I take it, that the property of-being-identical-with-a is a legitimate property, and a property that a actually has.

The real reason for not accepting the proposed definition of ‘identity’ is that the definition is circular. We are already presupposing identity in that the sufficient condition for identity contains the “identity-property”. This is denied by Baruch Brody when he claims that our definition is only impredicative, and not circular. Brody claims that a circular definition must have its definiendum as a part of the definiens, but, as he points out, ‘identity’ is not part of the definiens which is ‘the relation that holds between x and y just in case they have all of their properties in common’. Since the definiens only includes the general terms of ‘relation’ and ‘sharing a property in common’ Brody thinks it should be


28 Though I agree that our ordinary conception of properties is that of “plural instantiation”, such as the properties red, round, and tall.

29 Brody, Identity and Essence, p. 12.

30 Ibid. p.12.
possible to acquire, or grasp the meaning of ‘identity’ through the definition, i.e. that ‘identity’ is not a “conceptual primitive” as I have claimed.\(^{31}\)

But, is it really possible to grasp the definiens of the definition through those general terms without a pre-knowledge, or pre-understanding of identity? I do not think so. Firstly, what we need to understand is the idea of the objects \(x\) and \(y\) sharing all of their properties, including the "identity-property". But, this means that we must understand what relation identity is before the definition is given. Furthermore, the whole idea of having all properties in common, or of being indiscernible, requires a pre-understanding of identity, since, as E. J. Lowe has pointed out, ‘having all properties in common’ cannot mean anything other than \(x\) and \(y\) "having the same properties".\(^{32}\) This means that understanding the general term ‘sharing a property in common’ presupposes a prior understanding of identity, since we must be able to identify or distinguish properties from each other.\(^{33}\) We must, then, conceptually presuppose identity if the Identity of Indiscernibles should be a possible sufficient condition of identity.\(^{34}\)

Beside these formal conditions of identity, I think there are two metaphilosophical principles of identity which influence our understanding of identity. The metaphilosophical principles I am thinking of are (i) the intrinsicality of identity which determines that the identity between \(x\) and \(y\) should depend on


\(^{32}\) Lowe, *Kinds of Being*, p. 23.

\(^{33}\) Ibid. p. 23.

\(^{34}\) Whether the Identity of Indiscernibles really is a sufficient condition for identity is another problem. This depends upon what kinds of properties we include in the scope of the quantifier. If we accept, as Brody does, identity-properties and other relational properties in the set of properties, then we must conclude that the identity of Indiscernibles is a valid principle. But, if we qualify our condition with the additional proviso that we only allow "internal properties" of the objects \(a\) and \(b\) into the properties included in the scope of the quantifier of the Identity of Indiscernibles, then we can no longer claim the validity of the principle. But, regardless of the correctness or not of the principle of the Identity of Indiscernibles, I still think we can find its proper place in a theory of identity. What I would like to claim is that the principle, failing to provide a conceptual definition of identity, still has a role to play in our theory of identity. The principle has mistakenly been taken as a semantic-cum-metaphysical principle saying what identity consists in, when it really is an epistemological principle or inference taking us from similarity to identity. This means that, when we have objects \(a\) and \(b\) sharing the same properties, we infer that \(a\) and \(b\) are identical. If we have "identity-properties", or other properties that cannot be shared by two objects in the set of properties our inference would become a deductive inference, and the claim of identity would become deductively valid. But, if we do not have "identity-properties", or relational properties in the set, then we would have an inductive inference, and the claim of identity would not be guaranteed to be valid. This last case would be the most frequent case, since in most cases we only have knowledge of some of the common properties of \(a\) and \(b\), and usually do not know whether the property being-identical-to-\(a\) also is a property of \(b\).
nothing but the relation between \( x \) and \( y \), and (ii) the **determinateness of identity** which determines that it is always the case that either \( a \) is identical with \( b \) or \( a \) is not identical with \( b \). The intrinsicality of identity will be discussed in the chapter on Physical Criterianism, while the determinateness of identity will be discussed in the chapter on Non-Criterianism.

### 2.3 Numerical and qualitative identity

Traditionally it has been claimed that there is a distinction between, on the one hand, *numerical identity*, and on the other hand, *qualitative identity*. *Numerical identity*, it is claimed, is the universally applicable relation in which every object stands to itself, and to no other object than itself. *Qualitative identity*, on the other hand, is the relation of 'identity of qualities' in which a certain object \( a \) can be identical not only with itself, but also with a distinct object \( b \). If two objects are *exactly alike*, or *perfectly similar*, in all their internal (or non-relational) qualities, they are also qualitative identical.

That this distinction between *numerical identity* and *qualitative identity* is legitimate, and one that we usually can and do apply, can easily be apprehended if we consider a case where we are prepared to think that something has remained identical over time. Imagine that we have two balls on our billiard-table. Let us call them 'a' and 'b'. Suppose now, somewhat idealised, that \( a \) and \( b \) have exactly the same qualities. If this idealised situation occurred, then \( a \) and \( b \) would be qualitative identical. But, \( a \) and \( b \) are not numerically identical, since they are two different objects. We can, for instance, take \( a \) and paint it red. In that case, we still have two qualitatively distinct balls on our billiard-table.

---

35 I call these principles 'metaphilosophical principles' because they, contrary to the formal conditions of 'identity', are prior to any determination of the formal aspects. They are 'principles', rather than conditions, because the constraints they lay upon the relation of identity concern the framework behind identity, and not upon identity itself.

36 Besides this distinction between *numerical identity* and *qualitative identity* it is also often claimed that we can be using 'identity' in its 'gen-identical sense'. See for instance David Wiggins discussion of this notion in *Sameness and Substance*, pp 23-35, and pp 193-199. Roughly, 'gen-identity', or 'the unity-relation' as John Perry calls it, is the relation which combines, or unites different, but qualitatively similar parts of an object \( o \), and it is in virtue of this relation between the parts of \( o \), that all of those parts are parts of \( o \). (See for instance John Perry, "The Same F", *Philosophical Review*, 80, 1970.) Since this talk about 'gen-identity', or 'unity-relation', seems to be closely connected with interpreting objects as being four-dimensional, I will postpone the analysis of these notions, and consider them in connection with the four-dimensional interpretation of objects.
Another example, and one that does not involve identity over time, is that one can be the owner of two examples of the same edition of the book "Winnie-the-Pooh". In such a case we would be prepared to say that the two books are qualitatively identical, but not that they are one and the same object. That is, each individual example of the two books is numerically identical only to itself, but the two books are qualitatively identical with each other since they are exactly alike each other.

Though the distinction between numerical identity, and qualitative identity is both legitimate and common I think the distinction is formulated in a slippery terminology, since calling them both 'identity' seems to imply that we have different meanings of the notion 'identity'. But, I certainly oppose any such claims about 'identity' being an ambiguous notion. It is a mistake to understand 'numerical identity' and 'qualitative identity' as a distinction between different kinds, or senses of identity. When we say "Four is identical with two plus two", and "John is the same person as Dorothy" we do not mean different things with 'identity', or 'sameness'. What we mean is nothing but the unique relation that everything bears to itself, and only to itself, i.e. numerical identity.

There are two types of counter-objections to treating numerical identity as the only form of identity there is. Firstly, can we really mean numerical identity when we are talking about identity over time. Secondly, fission and fusion cases do not seem to be impossible for ordinary objects and hence 'personal identity' cannot mean numerical identity.37

### 2.4 Identity and change

Some philosophers, for instance David Hume, have thought the general idea of numerical identity over time to be incoherent.38 The general thesis that numerical identity over time is incoherent relies upon two different arguments. Both arguments take as their starting-point that our idea, or notion of 'identity over time' is that of an object that first exists at one time t, and then at another time t'.

---

37 Cases of fission and fusion are frequently discussed in the literature on personal identity over time, since it is one of the easiest ways to elucidate the more or less hidden intuitions we have concerning personal identity. See for instance Derek Parfit, Reasons and Persons, third edition, (Oxford: Oxford University Press, 1987). Though, cases of fission and fusion are common in the literature, there is little agreement how to understand these puzzling cases. Almost all commentators have their own conclusions regarding these cases.

Hume, for instance, said that our idea of identity over time is that "of an object, that remains invariable and uninterrupted thro’ a supposes’d variation of time. But, the argument goes, it is neither the case that an object can remain invariable, nor remain uninterrupted, during its persistence.

Firstly, objects that are viewed as existing over time are not invariable. For instance, consider a leaf on an apple-tree. In the summer the leaf is green, but in the autumn the leaf is red-and-yellow. Now, it is argued, it is a common fact that objects have different properties at different times, such that, at the time t the object O has the property E, and at t’ O has the completely different property F. But, according to the necessary conditions for identity, having different properties cannot be a consequence of a true identity-statement, and consequently we can conclude that we do not have one invariable object where there is a difference of properties. What this argument comes down to, then, is that the idea of change, being the same as variability of properties, is necessarily incompatible with something remaining the same over time.

The second argument against the possibility of identity over time denies the possibility of there being some object which is "uninterrupted thro’ a supposes’d variation of time" According to this argument, 'identity over time' is not only incompatible with change, but actually a contradiction in terms, since time is nothing but the experience of a succession of distinct objects. What we perceive, according to this line of thought, is several momentary and distinct objects which we mistakenly take for one single object. But, the important fact to notice is that it is only from the manner in which these momentary objects present themselves to us that we can form the idea of time. Time has no existence in isolation from the succession of interrupted objects. But, the argument continues, if time necessarily invokes a succession of experienced objects that are distinct from each other, and identity over time means that something remains the same through time, then identity over time just cannot be a coherent notion. What we have is either identity, in which case we never could form the idea of time, or we could form the idea of time from the experience of a multiplicity of objects in which case we only could form the idea of similarity over time. And, the argument continues, what we mistakenly take as identity is really, and cannot be anything other than degrees of similarity.

Neither of those arguments are, I think, convincing. The first thing we should notice is that both arguments depend upon the distinction between 'numerical'
and 'qualitative' identity, or as I prefer to call them, 'identity' and 'similarity'. But, the standard way of explaining the difference between numerical and qualitative identity is by looking at an object that exists over time. For instance, earlier we presented the distinction between numerical identity and qualitative identity by considering a case where a white ball was painted red, and we then said that a certain red ball is numerically identical with a white ball, though the white ball and the red ball are not qualitatively identical with each other. Admittedly, this cannot be considered as a decisive argument against the idea of the impossibility of numerical identity over time, since it is possible to explain the distinction without any objects being numerically the same over time. But, even though the fact that we often invoke numerical identity over time in our explanation of the distinction between numerical and qualitative identity, it cannot function as a conclusive argument against the idea of the impossibility of numerical identity over time. It gives us, though, a propensity for thinking numerical identity over time to be a coherent notion, and that something must be wrong with an argument showing that this is not so. But, then, what is wrong with these two arguments?

The first argument does not seriously consider the tensed character of the properties attributable of objects. Objects exists at different times. But, the properties of objects need not exist at different times. For instance, a certain object might be a child at one time, but an adult human being at another time. In this sense, properties are temporal. The instantiation of a property is always a relation between a certain object and a particular time. Consequently, there is nothing contradictory in assuming that one and the same object can at one time of its existence have the property P and at another time of its existence have the property not-P.

Concerning the second argument, what one first should notice, following P. F. Strawson, is that the possibility of even doubting that some objects remain the same over time is intelligible only if identity over time is possible. Suppose that the objection against the possibility of identity over time is correct. This would mean that what we experience is a sequence of momentary objects \(<x, x', y, z, ..., n,>\), where each momentary object is thought to be something completely independent of all other objects of the sequence. But now, if this idea of complete independence is correct, then a problem of identity over time would never occur. We would not only have a sequence of independent momentary objects \(<x, x', y, z, ..., n,>\), but also a sequence of independent momentary spatio-temporal sys-

---

tems <X, X', Y, Z,..., N> in which each individual object is experienced. And, doubting that one object x in the system X is identical with an object x' in a completely different system X' just does not seem to make any sense. There is nothing, given that the two objects belong to different spatio-temporal systems, which forces us to think the objects to be identical in the first place. In this sense, the intelligibility of doubting identity over time on the ground that objects pass through change presupposes that we have one single unified spatio-temporal system.

Furthermore, the second argument against the idea that numerical identity is possible rests upon the confusion of the content of a perception and the object of the perception. We need not assume, as for instance Hume did, that every perceptual content gives us a new object. In walking around a tree I have several different perceptions whose content differs from each other. But, this does not entail that I perceive several trees. I perceive one tree from different angles.

Furthermore, and this argument is directed towards both arguments against the possibility of there being identity over time, any change of objects in such a unified spatio-temporal system presupposes the identity over time of some kinds of objects of that system. Let us suppose that we have three objects x, y and z, where we say that x has changed into y. Let us further suppose that x is at place p at time t, and y is at place p' at t', and z is at place p at t'. In accounting for the fact that x has changed into y we claim that x has changed its place relative to z. But now, since z occupies the place p at t' must we not claim that z is at the very same place that x used to be, i.e. that the possibility of giving an account of change of objects presupposes that the identity over time of places is a fixed constant. More generally, the possibility of giving an account of change seems to presuppose that something is remained fixed over time. If we do not have some form of identity over time, then there would be no such thing as change. It is, in this sense, inconceivable to have change without also having something remaining the same.

Someone might argue that the identity over time of places is, as I have claimed, a necessary condition for change, but that this dependence of identity over time of places for discerning change only establishes that we must presuppose the existence of one single system of reference in which change can take place. It, the argument continues, says nothing about the identity over time of objects of that system of reference. Furthermore, it might be argued that it is from an epistemic point of view that we have to presuppose the identity over time of places. Presupposing identity over time of places is necessary for giving an account of change from an epistemic point of view, but this does not imply
that places are identical over time. Places, being the corner-stones of our system of reference, remain fixed over time, but this does not imply that places, nor anything else, must remain the same over time.

As to the first argument, it is true that I have established only that the system of reference must contain something fixed over time. But, I presume, this is enough. The original argument against the idea that something can remain identical over time is that this thought is logically incoherent. By showing that the system of reference presupposes, at least, the numerical identity over time of places I have established that the idea of numerical identity over time is logically coherent.

As to the second argument, it is true that we, from an epistemological point of view, must keep something fixed over time to be able to discern a change, but it is equally true that the notion of 'change' is conceptually tied to sameness over time. To begin with, it is reasonable to suppose that our conception of time is conceptually associated with some form of succession of states of the world. In this I am in absolute agreement with the Humean tradition. In conceiving time, we conceive the occurrence of happenings and events. Events are changes in the sense that an event is the transformation of a certain existing state into a new existing state. In this sense, time essentially involves the occurrence of change where the occurrence of a change consists in the beginning to be the case of something which previously was not the case. Time is, then, conceptually tied to the concept change in that if there were no changes, there would be no time.

Further, a change requires the existence of something at the time of the change which also existed prior to the change and this should be understood as an object which persists through the change. Supposing otherwise would imply that "everything in existence at the time of the supposed change has begun to exist at that very time" which would be the same as maintaining that that time is the beginning of time. But, the beginning of time is not the time of any change, since we at that time "cannot meaningfully speak of anything's beginning to be the case which was previously not the case". The notion of 'numerical identity over time' does not, then, in and by itself imply that there are different kinds, or senses of 'identity'.

43 Ibid. p. 75.
44 Ibid. p. 75.
45 Ibid. p. 75.
2.5 Cases of fission and fusion

Cases of fission and fusion are imaginable for almost all kinds of objects. For instance, a fusion of two cars is accomplished through the putting together into a car an equal amount of parts from two different cars, or a fission of a ship is brought about through gradually replacing all the parts of a ship, while we at the same time put those old parts together at another place. But, for the sake of brevity, I will only consider a case about a person, and moreover, only consider a case where this person, so to speak, divides.

Imagine that you are about to undergo a brain-transplantation. The reason for undergoing this extremely difficult and unpleasant operation is that you have been found to have a deadly disease. Let us now suppose that several brain-transplants have been performed during the years, but, unfortunately the result of those operations is distressing, since only fifty percent of them have been successful. But, luckily enough, it has been found out that several persons have survived with only one of the hemispheres intact, so the transplantation will be carried out by cutting of the Corpus Callosum connecting your hemispheres, and then placing each hemisphere in separate bodies. This would, it seems, increase your chance of survival of the operation. The operation takes place, and as a matter of fact, both of the transplanted hemispheres function satisfactorily. Let us suppose, for the sake of argument, that both persons closely resemble you psychologically. Calling the pre-operational person ‘a’, the post-operational person with the right hemisphere for ‘b’ and the post-operational person with the left hemisphere for ‘c’, it can be argued that this case really gives us an example which shows that ‘personal identity’ does not mean numerical identity and that we therefore have different senses of ‘identity’. The reason is very simple.

Firstly, suppose neither of the transplantations to be successful, then, of course, there would be no post-operational person identical with a. Secondly, suppose both of the transplantations to be successful, then neither of the post-operational persons are identical with a, since that would violate the transitivity of identity. Thirdly, suppose that only one of the transplantations succeeded, while the other was a failure. Under those circumstances it seems that we actually do have personal identity between a and b, or a and c. The only difference between the second case, where we have only one hemisphere successfully transplanted, and the case of fission is that in a case of fission we have two positive outcomes of the operation. There are no internal differences between a and b, or a and c in the two cases. And, as Derek Parfit once said, “How can a dou-
ble success be a failure?".46 Now, since it seems implausible that a double suc-
cess should be as bad as death, some philosophers have thought that the most
plausible way to describe a case of fission is by claiming that there really is per-
sonal identity between a and b, as well as between a and c.47 But, now, is it not
obvious, it is claimed, that ‘personal identity’ cannot mean numerical identity,
since the condition of transitivity of numerical identity is not satisfied. Even
though it is true that b is “personally identical” with a, and c is “personally
identical” with a, b is not “personally identical” with c. A similar kind of rea-
soning would show that cases of fusion also violate the transitivity of identity.
Since such cases of fission and fusion of different types of objects are easily
imaginable it simply does not make sense to claim that we always mean numeri-
cal identity by, for instance, ‘personal identity’ and ‘identity of car’. What we
mean in those cases is some other sense of ‘identity’, a sense that only has to
satisfy the conditions of reflexivity and symmetry.48

I do not think we need to be worried by the argument from the alleged cases
of fission and fusion, since these cases can be assimilated into a theory which
understands ‘identity’ as standing for numerical identity. What we need to do is
to take account of the difference between numerical identity and similarity. It is
not the case that we have personal identity between b and a, and personal iden-
tity between c and a. What we have is similarity, or even “personal similarity”
between b and a, and also between c and a. That b is similar to a at the same
time as the different object c also is similar to a is not a problem, since simi-
arity, contrary to identity, not is necessarily transitive.49

We ought, then, understand ‘identity’ in personal identity over time as stand-
ing for numerical identity. Claiming that a person existing at t2 is the same per-
son as the one existing at t1 should be interpreted as entailing that it is the very
same object which exists at t1 and at t2.

47 David Lewis has, for instance, claimed that a and b, as well as a and c are identical without
b and c being identical, on the ground that there exists a person-stage a” which is a stage of
both person b and person c. See David Lewis, ‘Survival and identity’, from The identities of
48 I have heard Jan Österberg defend such a theory at two different occasions. Both in
Uppsala on the 8/12-1995, and in Lund on the 15/12-1995.
49 ‘Qualitative identity’ is transitive in the case where it is interpreted as perfectly similar
with regard to internal qualities, but not in the case where it is interpreted as similar in some
respect Q.
2.6 Criterion of identity

The first thing to notice is that the term 'criterion' is ambiguous. There are at least three different possible uses of the term 'criterion', and it is only one of those uses that is of primary interest for the question of personal identity. If, for instance, someone says "The criterion used to select the control-group was that of age" then the usage of 'criterion' is that of meaning the same as the pragmatically structured principle, by the subject, of singling out a class of objects. This first sense of a 'criterion' we can call 'the pragmatic sense of criterion'. This sense of criterion is purely subject-oriented, not object-oriented. Its aim is not to tell anything of interest about the objects as being the very objects that they are, only how, or by what principle, the subject selected the group of objects from other objects.

The second sense of criterion is directly epistemological. In some contexts 'a criterion' is understood as meaning nothing but the epistemological evidence we have for the truth of a certain proposition. In this sense 'criterion' means nothing but the evidential-relation between the set of beliefs held by a person and the observations supporting or falsifying these beliefs. We use this second sense of 'criterion' when we ask for, or give, our evidence for believing in the identity between, for instance Hesperus and Phosphorus, or the red table I had yesterday and the green table I have today.

The first sense of 'criterion' has no connections at all to the notion of a criterion of identity, and need not be of any concern of ours. But, 'criterion' meaning the same as epistemological evidence has such a connection. If we claim that there exists a criterion of identity for persons by which we are able to determine whether a person P is identical with a person P', then it also seems to be required that we can evaluate the evidence we have for identifying P with P'. But, even though this evidential meaning of 'a criterion of identity' is related to the question of personal identity, it is not what primarily concerns us here. Remember that the second problem of personal identity was formulated thus, "In what, if anything, does personal identity consist?", and we can formulate a similar question about the identity between Hesperus and Phosphorus. But, interpreting 'criterion' as evidence can only tell us something about how we come to know that Hesperus and Phosphorus are identical and not what this identity consists in. Similarly, the evidence we have for the identity between a person P existing at the time t, and a person P' existing at another time t', cannot tell us what it is that makes P identical with P', only how we come to know that it is one and the same person. This means that there must be a third sense of 'criterion', a sense which purports to tell us something more than just how we come to know that
the identity-relation obtains. It must establish what the conditions there are that ground the identity.

Such a 'criterion of identity' for a sort S is a specification of a necessary and sufficient condition for objects x and y of the sort S to be (numerically) identical. We can write it of the form:

\[
\text{if } x \text{ and } y \text{ are objects of the sort } S, \text{ then } x = y \text{ iff } x \text{ and } y \text{ satisfy the condition } C_S(x, y),
\]

where \( C_S \) is some informative and non-circular specification of the relation 'x is the same S as y'. Taking an example from geometry, two directions x and y are identical if and only if, for any two straight lines L and L*, if L has direction x and L* has direction y, then L and L* are parallel.

This third sense of criterion I will call 'the semantic-cum-metaphysical sense of criterion'. It is a semantic determination, since it specifies the truth-conditions for identity statements of the form 'a = b', where 'a' and 'b' stand for objects of some sort S. And, it is a metaphysical determination, since it specifies necessary and sufficient conditions for an object x of sort S to be identical with an object y of the same sort.\(^{50}\)

But, it is important to notice that the semantic-cum-metaphysical understanding of 'a criterion of identity' in no way implies that a criterion of identity defines, or establishes necessary and sufficient conditions for 'identity' in general. A criterion of identity takes the notion of 'identity' as something given and uses it as a point of departure from which it determines semantic-cum-metaphysical identity-conditions for objects of a distinctive sort S.\(^{51}\) The identity-conditions are for a sort S the necessary and sufficient conditions that must obtain between objects x and y of sort S, if x is to be the same as y. The identity-conditions are dependent upon the sort or kind in question in the sense that different sorts or kinds of objects have different existence and persistence conditions.

This, though, does not mean that 'identity' is relative in the sense that "to conceive and judge of it [identity] aright, we must consider what idea the word it is applied to stands for (...)"\(^ {52}\) The idea that 'identity' is relative has been defended by P. T. Geach and it consists in the ideas that (i) it is a mistake to

---


think the complete coincidence of identicals to be a necessary consequence of
our notion of ‘identity’, (ii) “a particular x and a particular y can be the same F,
but not the same G, even though x is a G and y is a G”53 and (iii) expressions
involving absolute identity are either incomplete expressions, or “vague expres­sions of half-formed thoughts” that stand in need of being completed, or made
precise by a count noun.54 The idea that identity is absolute is expressed by:

The Absoluteness of Identity Thesis: if S and S’ are sorts, and x and y are
objects such that x is the same S as y and x and y are also of sort S*, then x
is the same S* as y.

This, of course, is the thesis that Geach is denying. In his view, judgements
about identity are always true or false relative to the description of the objects.
But, not every descriptional phrase we can come up with can function as a cor­rect classification of objects which could also lie as a foundation for relative
identity.55 The most suitable descriptional terms for describing objects in a rela­tive identity-judg ement are, according to Geach, “count nouns”,56 such as
‘person’, ‘man’, ‘tree’ and ‘computer’, but, he also thinks that “mass terms”,
such as ‘gold’ and ‘water’ function in this way.

Since Geach claims that every judgement about identity is true or false rela­tive to the count noun, or the mass term under which the objects fall, and these
count nouns, or mass terms are “not logical, nor even mathematical, term[s], it is
no business of logic or the philosophy of logic to supply a criterion of identity”
for objects falling under these terms.57 As Geach says:

even if it is the case that every man is a person and every person is a man, we still
cannot infer that the predicables “--- is the same person as---” and “--- is the same
man as---” coincide in application; this conclusion needs for its support an elabo­rate argument in the philosophy of mind, not a little logical manipulation in iden­
tity theory.58

That is, whether or not something a is always to coincide with b, cannot be
established through a logical inquiry into the conditions of identity, but only

53 John C. Nelson, Some comments on the identity conditions for material objects, A
Dissertation submitted to the Faculty of the Division of the Humanities, Department of
55 P. T. Geach, ‘Identity - a reply’, in Logic Matters, (Berkeley: University of California
56 Geach ‘Identity’, p. 3.
58 Geach, ‘Identity’, p. 11.
through a determination of what sort \( S \) of objects \( a \) and \( b \) are, and whether \( a \) and \( b \) coincide as objects of this sort \( S \), or not. Since some of the non-valid arguments in the relative language actually will become valid in the language of absolute identity we cannot analyse statements in the relative idiom as equivalent with the statement of absolute identity. The judgements "\( a \) is the same \( S \) as \( b \)" or "\( a \) is \( S \)-identical with \( b \)" are primitive and cannot be analysed into "\( a \) is an \( S \), and \( b \) is an \( S \), and \( a \) is identical with \( b \)" since the original expressions in the idiom of relative identity and the proposed equivalent expressions of absolute identity do not entail the same conclusions. The fundamental, or primitive form of an identity-judgement is always one which mentions a sort.

Even though I think the claim that identity always can be suited to an Idea is a plausible theory I do not think identity to be relative. As I understands it, it seems that a claim of relative identity already presupposes a particular use of 'identity' as being absolute.\(^5^9\) Let us suppose we pick out an object of the sort \( S \), and that we call it 'a'. As a matter of fact, \( a \) is actually the same \( S \) as the previously known object \( b \). But, we discover that \( a \) is not the same \( S' \) as \( b \), even though both \( a \) and \( b \) are \( S' \). Let us now introduce a new name for the object \( a \) which is an \( S' \), and call it 'a*'. As John C. Nelson has pointed out, the question we ought to direct to the advocates of the thesis of relative identity is by what right we are able to conclude that the object \( a \) which is an \( S \), is also an object which is an \( S' \), and the object \( a* \) which is an \( S' \) also is an object which is an \( S' \)?\(^6^0\) As far as I can see, the only way for an advocate of the thesis of relative identity to be able to establish this is through using the Identity of Indiscernibles. That is, \( a \), which is an \( S \), is also an \( S' \), and \( a* \), which is an \( S' \), is also an \( S \), only because \( a \) and \( a* \) occupies the same places at the same times, and that distinct objects cannot occupy the same places at the same times.\(^6^1\) And from the Indiscernibility of Identicals it then follows that every property had by \( a \), must also be a property of \( a* \), and vice versa.

Though this seems to be a promising manoeuvre, it cannot be maintained by an advocate of relative identity. Firstly, if we suppose the Identity of Indiscernibles to be correct and that the use of it is perfectly legitimate to establish that \( a \) is an \( S' \), and \( a* \) is an \( S \), it is still the case that it is possible to establish the identity of \( a \), which is an \( S \), with \( a \), which is an \( S' \), and \( a* \), which is an \( S' \), with

---

\(^{59}\) Nelson, *Some comments on the identity conditions for material objects*, p.107.

\(^{60}\) Ibid. p. 106.

\(^{61}\) Ibid. p. 107.
the a*, which is an S, only through using absolute identity.\textsuperscript{62} As I earlier claimed, the whole idea of having all properties in common requires that we can determine when we have the same property, and when we have different properties, since we then claimed that ‘having all properties in common’ means the same as “having the same properties”.\textsuperscript{63} But, since ‘same property’ means \textit{numerically the same} property, the idea of relative identity can only be explained by accepting that identity is not always relative.\textsuperscript{64}

Secondly, the reason for claiming that a is an S' as well as an S, and that a* is an S' as well as an S, was that they occupied the same place at the same time together with the principle that objects occupying the same place at the same time are identical. But, this last principle is not, I think, true as it stands. For instance, a table \textit{k} occupies the same place at the same time as the sum of atoms \textit{s}_1, \ldots, \textit{s}_n which composes the table. But this does not mean that we have to claim that the table exists and the atoms do not, or the atoms exist and the table does not, or that the table is identical with the atoms. It is the case that both the table exists and the atoms exist, but they are different kinds of things. They are different kind of things, since they have different persistence-conditions. Because they are different kinds of things, the table can occupy the same place at the same time as the atoms without being identical with the atoms. But, this means then, that identifying \textit{a} with \textit{a*} is not an easy matter. Since we know, that \textit{a} is an S, and we also know that \textit{a*} is an S', and we also know that S is not the same sort, or kind as S', then we have no ground to conclude that \textit{a} and \textit{a*} are the same from the mere fact that they have the same spatio-temporal relational properties. That is, \textit{a} can be one thing, and \textit{a*} a completely different thing, even though they share all their spatio-temporal locations.

Finally, in arguing for the possibility of establishing that it is the very same object \textit{a} which is both an S and an S', and that \textit{a} is in fact the same as \textit{a*}, the advocate of relative identity made use of the Indiscernibility of Identicals. It was claimed that, since \textit{a} is the same as \textit{a*}, then every property had by \textit{a}, must also be a property of \textit{a*}, and vice versa. Though there is nothing wrong with this principle, the advocate of relative identity cannot, as Nelson has pointed out, make any use of it.\textsuperscript{65} The principle says that \textit{a necessary consequence of \textit{x} being

\textsuperscript{62} \textit{Ibid.} p. 106.

\textsuperscript{63} Lowe, \textit{Kinds of Being}, p. 23.

\textsuperscript{64} If someone thinks that ‘same property’ does not mean numerical sameness of properties, then think that there is only one series of properties involved in the case, since the case is supposed to involve just one individual.

\textsuperscript{65} Nelson, \textit{Some comments on the identity conditions for material objects}, pp. 107-108.
identical with y, is that x and y should coincide, or have all of their properties in common. But now, this principle is not something which could be assumed by an advocate of relative identity, since relative identity is the claim that x and y can be identical as being objects of the sort S, while not being identical as being objects of the sort S'. Not only is the Indiscernibility of Identity incompatible with relative identity, the Indiscernibility of Identity is exactly what relative identity opposes in the thesis of absolute identity. But this means, that an advocate of relative identity cannot use this principle establishing that it really is one and the same object that is both an S and an S'.

We have, then, no reason to suppose the thesis of relativity of identity to be correct. Absolute identity is correct and identity-judgements involving a relativization to a particular sort such as 'a is the same S as b' should be understood as equivalent to 'a is an S, and b is an S, and a is the same as b'.
3. SORTAL INDIVIDUATION AND PERSONS

3.1 Introduction

In this chapter I will defend the sortal dependency of individuation, i.e., the thought that an unequivocal identification of a particular object rests upon classifying the object as falling under a sortal term $S$. The version of the thesis that I will defend has to a large extent been defended by both David Wiggins and E. J. Lowe, though at some points I diverge from both of their accounts. I will claim that this thesis should be understood as a conceptual and ontological thesis in which objects must be thought of as belonging to some specifiable sort $S$ if we are to be able to individuate the objects. I will claim that sortal terms are associated with sortal concepts, and that it is the sortal concepts which provide us with "principles" of determining how many objects of a certain sort $S$ there are in a certain region during a certain period of time. I will also claim that certain sortal concepts should be understood according to an externalistic account of meaning. Some sortal terms are *substance sortal terms*, i.e. they are by necessity applicable to an object during the object's entire existence, while other sortal terms are *phased sortal terms* and may only apply to an object during a shorter part of the object's existence. I will also claim that 'person' is a sortal term and that we have prima facie evidence for thinking it to be a substance sortal term whose meaning is determined by reference to human persons.

I will claim that sortal terms exist in a hierarchy and that certain sortal terms are ontologically primitive in relation to other sortal terms. This account of ontological priority of sortal terms is then used as a way of establishing that persons are objects that endure over time. The general idea of an enduring object is that

---


42 Wiggins, *Sameness and Substance*, and Lowe, *Kinds of Being*. It is also the case that Wiggins and Lowe to some extent have different accounts of the sortal dependency of individuation. But, the similarities between their accounts are more salient than their dissimilarities.
of an object which is "wholly present" at every time of its existence over time. Contrasted with this view is the perdurance view of objects in which an object's existence over time is thought to consist in a relation between different temporal parts of a four-dimensional object. What I claim is that it is a part of our concept of being a person that persons endure. That is, it is a part of our concept of being a person that a person is numerically the same at different times \( t \) and \( t' \) during its existence. This means, then, that it is a conceptual necessity that persons persist through change and time in accordance with the phenomenological experiences had by persons concerning their own and others' identity. This is "established" by showing that there are certain problems with the alternative perdurance view, and showing that the alternative perdurance view is ontologically dependent upon the endurance view of ourselves.

3.2 Sortal individuation, sortal terms and sortal concepts

The general idea of the thesis of the sortal dependency of individuation, and the one I will fall back on, goes back to P. F. Strawson and has been defended by David Wiggins and E. J. Lowe.\(^{43}\) According to this view, identifying an object as being a particular object is possible only if the object is simultaneously classified as falling under a sortal term associated with a sortal concept\(^{44}\) which "supplies a principle for distinguishing and counting the particulars to which it applies".\(^{45}\) This account of the sortal dependency of individuation entails that:

(i) every existing object \( x \) belongs to some sort \( S \),

(ii) for every existing object \( x \), there exists a sortal term ‘\( S \)’ which unequivocally identifies \( x \) as falling under ‘\( S \)’,


\(^{45}\) P. F. Strawson, *Individuals*, p. 168. Though claiming this, one need not demand that a sortal term must be associated with an explicit principle of enumeration. This idea is endorsed by Strawson himself in his later writings. See for instance P. F. Strawson, ‘Individuals’ in *Philosophical problems today*, Vol. 1, (Haag: Kluwer Academic Publishers, 1994), pp. 21-44. In this paper Strawson claims that there are certain forms of individuals for which no criterion of identity can be given, but to which we are nevertheless able to make identifiable singular reference. Such examples would be hair-styles, manners of speaking, jealousy. (see p 29). Similar ideas can be found in Wiggins, *Sameness and Substance*, p. 8, and Lowe, *Kinds of Being* p 1.
(iii) this sortal term ‘S’ provides a principle of individuation for objects of the sort S, i.e., there is a determinate answer to the question for any two objects $a$, $b$, of the sort $S$ whether $a$ is the same $S$ as $b$; and

(iv) if $a$, $b$ are objects of the sort $S$, then $a$ is identical to $b$ if and only if $a$ is the same $S$ as $b$.

To be able to individuate a particular object – to be able to have a thought of or experience a distinct object – it is required that the object is thought of or experienced as a “this such”\(^\text{46}\) which distinguishes it from objects in its surroundings.

This, though, does not imply that we always have to identify objects via sortal terms. The two central features of the sortal dependency of individuation are (i) the ontological thesis that every individuated object has to belong to some sort $S$, and (ii) the conceptual thesis that we cannot meaningfully speak of, nor think of, a particular object unless it rests upon an unequivocal identification via a conceptual content of a sortal term ‘$S$’. It does not entail the epistemological thesis that every individuated object must, in the act of identification, be classified as falling under a sortal term ‘$S$’.

It is a fact that we are able to demonstratively pick out objects without passing through a determination of any sortal classification of the object. For instance, very often we determine what object we have in mind by just associating with the object a demonstrative ‘that’, or ‘this’. This is especially obvious in cases where we encounter objects we have never seen before, but it also occurs in ordinary circumstances where we already know what sort the object we experience belongs to. In this sense, it is not the case that every identified object must be classified as falling under a sort.

But, the fact that we do not need to identify objects via some sortal at every single instance of identification, does not imply that the thesis of sortal dependency of individuation is false. The fact is that such non-sortal demonstrative individuation is a highly abstract way of individuation which requires, for its possibility, a more primitive act of individuation in which objects are sortally individuated. It is only because a person has carried out several acts of individuation through the use of sorts, and as a consequence of that determined the boundaries of ordinary objects of experience, that he later can accomplish a non-sortal individuation with the help of demonstratives such as ‘that $x$’ and ‘this $x$’. This is partly what is contained in the conceptual thesis of sortal individuation. A person has, so to speak, acquired a certain kind of sensitivity to how objects behave and what they look like with the help of sortal individuation, and once

\(^{46}\) Wiggins, *Sameness and Substance*, p. 6 and p 15.
learned, this knowledge of the behaviour and structure of objects can be used in new situations without invoking any sortals. That is, even though it is possible to individuate objects without sortals under certain circumstances, these acts of non-sortal individuation depend, conceptually, upon a prior familiarity with sortally individuated objects. A second, but nonetheless, important feature of the conceptual thesis of sortal individuation is the general acceptance of the conceptual assumption that every object belongs to a sort $S$ which it is possible to refer to with the constructed sortal term ‘$S$’. In this sense, it is conceptually assumed that the abstract and non-sortal individuation via direct demonstratives is always incomplete in the sense that a person using demonstrative individuation always can be asked the question "That what?", or "This what?". Individuation is, then, the simultaneous process of identification of a particular object $x$ and the classification of that particular object $x$ as falling under a sortal term through a sortal concept by which it is both generally possible to enumerate individual objects falling under the sortal term and to refer to the entities which fall under the sortal term.\footnote{Wiggins, *Sameness and Substance*, pp. 47-101, and Lowe, *Kinds of Being*, pp. 9-42.}

Thus, each sortal term $T$ is associated with a sortal concept which determines the sort designated by the term. We can think of the sortal concept $T$ as consisting of three components:

(i) one component, which we may call the criterion of application of $T$, determines the extension of the term $T$,

(ii) one component which, if non-circularly specified, would be the criterion of identity (or the principle of individuation) of $T$,

(iii) “a stereotype” which provides the sortal term with a descriptive content.

Though sortal concepts might provide a criterion of identity or principle of individuation for the sort $S$, this is not a necessary feature of sortal concepts. The reason for this is that there might be certain sortal terms that are “basic sortals”. This idea goes back to E. J. Lowe.\footnote{Other ideas entailed in a sortal term being a basic sortal are presented in chapter 9, section 9.2.} One of the central ideas of a sortal term being a basic sortal is that the term does not provide us with a non-circular and informative criterion of identity for objects falling under the term.\footnote{Wiggins, *Sameness and Substance*, pp. 47-101, and Lowe, *Kinds of Being*, pp. 9-42.} In this sense,
a sortal term 'S' is a basic sortal term if and only if the identity over time of objects falling under 'S' cannot be understood without already employing or presupposing the concept of S-identity.

One of those sortal terms that is basic is, I think, 'person'. We are able to individuate and enumerate objects that are persons, even though we lack a criterion of identity for persons.

The general idea of an account of sortal terms and sortal concepts is that the sortal concepts associated with sortal terms should provide us with a principled way of determining whether an object falls under a certain sortal term or not, and how many objects there are at a certain place during a certain period of time. Examples of sortal terms are 'table', 'baker', 'human being' and 'person'. All of these terms are individuating terms in that they pick out individual objects. For instance, the associated conceptual content of the sortal term 'cat' should determine whether the experienced white cat in my room at the time t is distinct or identical with the experienced white cat in my room at the earlier time t₁. In this sense, sortal concepts associated with sortal terms should contain some content by which it could be determined how objects falling under the sortal terms persist and are re-identified. That is, the conceptual principle associated with 'S' should determine what it is to be an S where this consists in (i) what individuality as an S is, and (ii) what kind of continuity the continuing existence of an individual S requires. Both mass terms, such as 'gold' and 'water', and dummy sortal terms, such as 'object' and 'thing' lack the required association with a determinable principle for counting numerically distinct objects falling under the terms and are therefore not sortal terms. In this sense, a term 'S' is a sortal term if it is a general noun of which it is intelligible to ask how many objects falling under the general noun there are, or have been, at a certain region of space during a certain period of time. The intelligibility of the how many-question is accounted for by the sortal concept associated with the term. That is, the how many-question should not only be intelligibly asked concerning the objects existing at a particular time and place, but also concerning the number of objects at a place during a shorter or longer period of time.

The way one should understand sortal terms is along the lines of the so-called 'New Theory of Reference' developed by Saul Kripke, Hilary Putnam, David

49 Claiming that 'person' is a basic sortal is a strong claim, but not something I will argue for in this chapter. It will be done when we undertake an investigation of different Criterianist theories concerning personal identity over time, and we will see that a criterion of identity over time cannot be non-circularly and informatively specified.
Kaplan among others. The general idea of this account, especially as it is formulated by Putnam with regard to natural kind terms, is that:

(i) natural kind terms resemble indexical terms in that they are not purely descriptional,

(ii) the meaning of a natural kind term involves an ostensive element,

(iii) an object falls under the extension of a natural kind term 'K' (or, is of the kind K), in any possible world, iff it has the property of being the same as this, where 'this' rigidly picks out a sample falling under the natural kind term in the actual world,

(iv) a natural kind term is associated with a conceptual stereotype consisting of general features paradigmatically possessed by instances falling under the term. That an object resembles the stereotype is a defeasible indication that it belongs to the kind.

I will proceed on the assumption that this account of the meaning of natural kind terms is applicable to (at least some fundamental) sortal terms. Thus, the sortal concept associated with a sortal term determines: (a) a stereotype providing the term with a descriptive content, (b) the extension of the term in every world, (c) whether two objects belonging to the extension are numerically the same or not. (The reader should remember that clause (b) and (c) above would yield a criterion of application and a criterion of identity if they could be non-circularly specified.)

The idea here is that, just as in cases of natural kind terms, it is through grasping the stereotype associated with a sortal term 'S' that we normally do identify objects as falling under 'S'. But, the stereotype only provides a defeasible condition for falling under the sortal term. The central feature of a (fundamental) sortal term 'S' is the ostensively defined association with a particular object in the actual world to which every object must have the same relation to if they are to fall under the sortal term 'S'.

I want to emphasise two important consequences of accepting this externalistic account of sortal concepts. The first is that it implies that the conceptual analysis which we favour cannot be purely conceptual, or \textit{a priori.} Though containing certain elements of \textit{a priori} conceptual analysis, the analysis we favour


must also be sensitive to empirical factors. Since the content of a sortal term contains an external constituent we have to incorporate empirical findings into our account of what it is to be an object of a certain sort S. That is, since the meaning of the term is fixed with the help of the sameness relation to an actual object as the paradigm, what is essential for being that particular object will also be essential for being an object of the sort S.

The second consequence of the externalistic account of, at least, some sortal terms and sortal concepts is that the fundamental sortal terms should be understood as "rigid designators" and that a sortal term 'S' picks out the same internal structure of objects falling under the term in every possible world in which there exists a sort S to which 'S' refers. As a consequence of this, it is metaphysically impossible that an individual object could fall under a certain sortal 'S', but not have the same internal structure as objects falling under the sortal term in the actual world. For instance, 'tiger' is a sortal term which is determined in relation to actual tigers. Because of this, any entity which is a tiger must possess the same internal biological structure as actual tigers. It is, then, an essential condition for an object being a tiger, that that very object has a biological structure similar to actual tigers of the world.

### 3.3 Substance sortal terms and phased sortal terms

The primary function of sortal terms is to individuate objects through classifying them as being "this such". Doing this is to give an answer to the what is it-question. But, since one and the same object can be individuated via different sortal concepts, not all sortal terms can be understood as equally good candidates for determining the what is it-question. For instance, we can individuate a as a foal, a as a grendling or a as a horse.

I will claim that the best candidates for answering the what is it-question in relation to the individuation of objects are substance sortal terms such as 'table', 'human being', 'frog' and 'horse'. A substance sortal term individuates an object through the longest prolongation of the object and is because of this semantically simple in the sense that its true attribution to an individual object is not dependent on whether other sortal terms are attributable to it. Contrasted

---

52 Saul Kripke, Naming and Necessity, p. 48
54 Lowe, Kinds of Being, pp. 30-31.
with substance sortal terms are phased sortal terms which may be applicable of an object only during a part of the object's whole existence. Phased sortal terms are semantically complex in that they are adjectival in relation to substance sortal terms. The only way of grasping the meaning of a phased sortal term is by simultaneously grasping some substance sortal term which determines part of the meaning of the phased sortal term.

A reason for thinking that a substance sortal term is particularly suited for answering the what is it-question is that such a term is associated with a sortal concept entailing essential properties of objects falling under the sortal term. Since substance sortal terms individuate objects through their entire existence, substance sortal terms determine essential properties of objects falling under them. If an object \( a \) is individuated under the substance sortal term ‘S’, then \( a \) cannot cease to be an S without thereby also ceasing to exist. The associated concept of a substance sortal term is compatible with, and provides an account of, those alterations an object can go through without ceasing to exist. For instance, to understand, and to be able to individuate something as falling under ‘infant’ one must have some conception of the object as being a living, biological organism, i.e., an object which could be individuated as an animal. A sortal term which allows more radical alterations of individual objects falling under it is ‘frog’. In the early part of the life of an object to which we apply the sortal term ‘frog’ nothing of the object’s morphological structure seems to resemble the paradigmatic case of a frog. Still, it is a part of the conceptual content associated with the sortal term ‘frog’ to have been a tadpole, since every grown up frog has been a tadpole at some time during its existence. The sortal term ‘tadpole’, on the other hand, cannot account for this radical alteration of the object, since ‘tadpole’ can be truly applied to an object only during some shorter interval of that object’s existence. An interesting fact, though not something I will pursue any further, about the phased sortal terms dependence upon substance sortal terms is that some phased sortals must be considered as necessary restrictions of the substance-sortal, while others can be understood as only contingently so. For instance, it is a necessary condition for an animal to pass through a time of infancy, so that ‘infant’ should be understood as a necessary phased sortal of the sortal term ‘animal’. A completely contingent phased sortal would, for example, be ‘baker’, ‘alcoholic’, and ‘teacher’ as regards the substance sortal ‘human being’.

---

3.4 Identity, sortal individuation and criteria of identity

Given the distinction between substance sortal terms that refer to substance sorts and phased sortal terms that refer to phased sorts, I interpret the thesis of the sortal dependency of individuation as implying the following principles:

The Sortal Individuation Thesis: every object $x$ belongs to some substance sort $S$.

The Specification Thesis: for any substance sort $S$ and any objects $x$ and $y$ whatsoever, $x$ is the same $S$ as $y$ (in short, $x=Sy$) iff $x$ and $y$ are of sort $S$ and $x=y$.

From these conditions it is possible to infer:

The Sortal Dependency of Identity: for any objects $x$ and $y$, $x=y$ iff there exists a substance sort $S$ such that $x$ is the same $S$ as $y$,

and

The Absoluteness of Identity: if $S$ and $S^*$ are substance sorts and $x$ and $y$ are objects such that $x$ is the same $S$ as $y$, and $x$ and $y$ are of sort $S^*$, then $x$ is the same $S^*$ as $y$.

To give a criterion of identity for objects of a substance sort $S$, then, is to give an informative and non-circular specification of the relation being the same $S$ as. It is possible, though, to distinguish between a weaker and a stronger notion of identity-criterion. The weaker kind of identity-criterion is of the form:

If $x$ and $y$ are objects of the sort $S$, then $x=Sy$ if and only if $C_S(x, y)$.

The stronger form of identity-criterion is of the form:

For all objects $x$ and $y$, $x=Sy$ if and only if $C_S(x, y)$,

where $C_S(x, y)$ is a condition that is false if $x$ or $y$ is not an object of the sort $S$.

The existence of a strong identity-criterion for objects of a sort $S$ implies that there also exists a weak identity-criterion for objects of the sort $S$, but not the converse. The strong form of identity criterion allows us to eliminate talk of the relation $=S$ in favour of talk of the relation $C_S$ in all contexts, while the weak form of identity criterion only makes it possible to eliminate '$=S$' in favour of $C_S$ in some contexts. For example, if 'a' is the name of a person and 'b' is the name of something that is not a person, it is not possible to eliminate the statement 'it is not the case that $a=Sb$' by specifying a weak criterion of identity for objects of the sort $S$. Given a strong criterion of identity for objects of the sort $S$, though, it
is possible to eliminate 'it is not the case that \( a =_S b \)' in favour of 'it is not the case that \( C_S(a, b) \).

Since \( x \) is of sort \( S \) if and only if \( x =_S x \), it is possible to infer, given a strong criterion of identity for objects of the sort \( S \), a criterion of application for objects of the sort \( S \) of the form:

\[
x \text{ is of sort } S \text{ if and only if } C_S(x, x).
\]

Thus, a strong criterion of identity for the sort \( S \) is equivalent to a specification of what it is to be an \( S \), in the strong sense of specifying both a necessary and sufficient condition for belonging to the sort \( S \) and a weak criterion of identity for objects of sort \( S \).

In what follows I will always mean by 'a criterion of identity for objects of sort \( S \)' the strong form of identity-criterion (except when otherwise explicitly mentioned). This is because, as already mentioned, questions concerning the identity of objects of some sort \( S \) is one way of determining what it is to be an object of the sort \( S \). That is, a criterion of identity should determine under what conditions any objects \( x \) and \( y \) are \( S \)-identical and this requires a specification of both a criterion of application for objects of the sort \( S \) and a criterion of identity for objects of the sort \( S \).

3.5 The sortal term ‘person’

It seems obvious that ‘person’ is a sortal term. ‘Person’ is a term which seems to individuate and enumerate objects in the world. It is a general count noun of which it is possible, at least, in ordinary cases, to give determinate answers to the *how many* question. In all everyday circumstances we are able to count how many persons we have in a particular place at a time. Secondly, not only is it possible to give a determinate answer to the *how many* question on specific occasions, it always seems intelligible to ask the *how many* persons there are at a certain place. Regardless of what account of persons and personal identity one is advocating there will never occur a situation where it would not be meaningful to ask *how many instances there are of the term* ‘person’ at a particular place during a certain time. But, this means that ‘person’ is a sortal term.

Whether ‘person’ is a substance sortal term or a phased sortal term is not something I will be able to establish at this stage. At this stage I just want to argue for the *prima facie* plausibility of understanding ‘person’ as a substance sortal term. When considering the notion ‘person’ we are inclined to understand the coming into being of a person as the coming into being of an object and that
the ceasing of being of a person as the ceasing of being of an object. For instance, if you have a child which dies in a tragic car-accident, you would consider this as the elimination of one object from the world. But, since your child *prima facie* is a person, this indicates that the death of the person that is your child is the elimination of an particular object. The same is true, at least *prima facie*, of giving birth to a child. Compare these cases with the ceasing of being a baker. You have a child that is a baker. When he entered into bakery you did not treat this as the addition of a new object, a baker, into the world. The same is true if he quits being a baker. You do not understand ceasing to be as a baker as the ceasing to be of a particular object. Your child may continue to exist even though he quits baking. But, since 'baker' is a phased sortal this seems to indicate that 'person' must be of a different category than phased sortal terms. It is, then, *prima facie* a substance sortal term in that *being-a-person* seems to be a property which an object cannot loose without ceasing to be.

Whether 'person' is a rigid designator referring to one and the same sort in every possible world cannot be established at this stage. My aim is that the entire thesis, looked upon as a closely integrated account of persons and personal identity, should provide the reader with an argument, though probably not conclusive, for thinking 'person' to be a rigid designator which, in every possible world, refers to the sort of objects it refers to in the actual world. But, as in the case of understanding 'person' as a substance sortal term I do think there is a *prima facie* indication that 'person' is one of those fundamental sortal terms whose associated content is externally determined through an ostensive definition to an actual instance of the world.

As far as we know, there are at least one group of individuals who definitively belong to the sort *person*. One may question whether God, Angels, dolphins, chimpanzees, Martians and even computers should be thought of as belonging to the sort *person*. But, when it comes to the question whether you and I, who are ordinary human persons, qualify as being persons there simply cannot exist any doubt. Ordinary human persons are, in this sense, the paradigm of what it is to be a person. These are the persons that we actually can be certain that we have encountered, and as such they provide us with the determining features necessarily possessed by any person, even if it were God, Angels, dolphins, chimpanzees, Martians or computers. This latter group of entities could only belong to the sort *person* in virtue of being in some important aspects the same as ordinary human persons such as you and I. They must, if they should fall under the extension of 'person', have the same _p_ relation to this person, where 'this' refers to a human person in the world. So far, though, it does not say what kind of
features of the actual individuals are to be necessary for falling under ‘person’. For instance, it might be the case that ‘person’ is defined in virtue of possessing the same, or very similar, mental capacities as ordinary human persons, or it might be that our concept person is defined in relation to some form of deeper level structure discernible in ordinary human persons. Regardless of what the exact conditions for personhood are determined to be it is still the case that they are defined in relation to certain conditions discernible in actual human persons and this gives us a prima facie reason to suppose that ‘person’ is a rigid designator which picks out one and the same sort in every possible world.

3.6 Ontological priority

It is reasonable to assume that objects belonging to certain substance sorts are ontologically prior, or primitive to objects of others sorts. This idea goes back to P. F. Strawson’s thesis of ontological priority. The general idea of Strawson’s thesis is that every conceptual scheme, or conceptual structure, in which the world is given to a sentient being contains individuals of different levels and that some individuals seem to function as necessary foundations for the existence of other individuals in the sense that the latter individuals are identification-dependent upon the former. The idea, then, is that it is a necessary requirement for being able to conceive of or think about an individual of a higher level that, implicitly or explicitly, the identification of some other, and more basic, individuals are made.

Since individuation is sortal dependent this idea of ontological priority can be stated in terms of the individuation of objects via the use of sortal terms. That is, if there is a sort of object β, such that we cannot individuate objects of sort β without reference to objects of another sort, α, but we can individuate objects of sort α without reference to objects of sort β, then it would be the case that we can talk about β-objects depends upon the ability to talk about α-objects, but not the other way around.

The idea is that there is a more fundamental, or basic level of description at which objects can be presented to us, and that this level of presentation of objects is ontologically primitive in relation to some other description of the

57 Strawson, Individuals, p. 15.
58 Ibid. p. 15. Strawson’s term is ‘particular’.
59 Ibid. p. 15.
objects. That is, we can as a matter of fact, give descriptions of the existence of objects without describing this fundamental level of objects underlying the other objects, but this can only be done through presupposing the more fundamental ontological level. In this sense, then, what is ontologically prior cannot be totally analysed or defined in terms of what is ontologically secondary. For instance, if there is a sort of objects $\alpha$ which are ontologically prior to a sort of objects $\beta$, then an analysis or definition of objects of the sort $\alpha$ cannot be done by using objects of the sort $\beta$, and we can say that objects of the sort $\alpha$ are ontologically prior to objects of the sort $\beta$. In individuating a particular object via classifying the object as being of a certain sort, we not only classify the object itself, we also specify what position the object has in our scheme of thought. Certain kinds of sortal terms contain a content which would, in a conceptual analysis, show that the use of the sortal concept as an individuating device in a particular case depends upon some more basic, fundamental facts.

Someone might argue that using the terms 'ontological priority', or 'ontological primitive' is misrepresenting the issue at stake. In Strawson's account of ontological priority, it could be argued, that what Strawson investigated was the conceptual scheme that is presupposed in our use of ordinary language and thinking about the world, and as such, his investigation can, at best, establish dependencies between different concepts. That is, the fact that in our conceptual scheme we attribute primitiveness to objects of the sort $\alpha$ relative to that of $\beta$, does not imply anything about the ontological primitiveness of $\alpha$ to that of $\beta$'s. What it at best can establish is that our concept $\alpha$ is primitive to our concept $\beta$. I do agree that Strawson's account cannot, in and by itself, take us any further than to an analysis of conceptual priorities between different concepts.

But, the account of sortal concepts I am favouring is externalistic in the sense that we take it that the meaning of a sortal term is given by a concept, or a sense which is constituted by an ostensive definition to an actual object. Because I have this external, or world dependent feature as a part of my associated content of a sortal term, it is plausible to uphold a position which consists of the claim that the analysis of dependencies between concepts of our world view establishes ontological priorities in the truest sense of 'ontological'.
3.7 The persistence of persons: the Perdurance view

The problem of how persons persist is a part of the more general Heraclitean problem of how any ordinary object o of any kind K can at one time t have the property F, and at another time t' have the property F' (where F is not identical to F'), and still remain one and the same object o of that kind K. According to the endurance view objects persist in virtue of being wholly present at every single time of their existence, and according to the perdurance view objects persist in virtue of the existence of a sequence of related, but non-identical, temporal segments of a person. The general intuitive idea of objects as perduring is that the object’s existence over time consists in a relation between different temporal parts of a four-dimensional object, while the general idea of an enduring object is that is a “continuant”, or an object which is “wholly present” at every time of its existence over time. Philosophers who have been advocating a perdurance view are David Hume, W. V. O. Quine, David Lewis and John Perry. Philosophers who have been advocating an endurance view, on the other hand, are for instance P. F. Strawson, and David Wiggins.

The general approach of the perdurance view is that conceptual oddities involving attribution of identity over time to a changing object can be resolved by accepting that our ordinary language is ambiguous in its referring expressions and that the puzzlement is a consequence of us not keeping track of what sort of entity a certain referring expression is supposed to be referring to at a certain time.

According to the perdurance view ordinary persisting objects, such as cars and horses are extended processes consisting of three-dimensional temporal parts. The reason for this is that time is understood as an additional dimension and that objects are the entities that fill a region of space-time. This understanding of ordinary objects implies that ordinary objects are aggregates whose composing parts are temporal parts of the objects. The temporal part-whole relation of persisting objects is understood in analogy with the spatial


61 See, for instance, Strawson, Individuals and Wiggins, Sameness and Substance.


part-whole relation of ordinary physical objects as they are at a certain time $t$. For instance, a chair $C$ existing at the time $t$ can be said to be composed of the spatial parts; the seat of $C$, the back of $C$ and the legs of $C$. In the same way, the chair $C$ persisting from $t_1$ until $t_3$ consists of the temporal parts; the chair $C$ at $t_1$, the chair $C$ at $t_2$ and the chair $C$ at $t_3$.\(^{64}\)

A temporal part of an object $O$ is not to be identified with the object $O$, but is a certain filled region of space-time which at a certain time $t$ constitutes the object $O$. A temporal part of an object $O$ is an object such that the temporal part of $O$ occupies all of the space occupied by $O$ at time $t$, but not necessarily all the space occupied by $O$.\(^{65}\) This description of temporal parts is compatible with several different understandings of the nature of temporal parts. If the temporal parts are understood as persisting over time they could themselves be understood as processes, events, or careers of objects. In that case we would have an extended time-sequence of an object. If temporal parts are understood as momentary objects they could be taken as the way things are at a time, or a space-point. In that case we would have a momentary time-slice of an object. The important thing with temporal parts is that they should be understood as the constituting part of the persisting object they are the very parts of, and that they should fill some sub-region of space-time of the region that the whole object occupies. In this sense, it is in virtue of the existence of an relation between the time-related parts of a persisting object that a persisting object exists. For instance, the temporal parts of the chair are the constituent parts of the persisting chair, and each temporal part exists at different times of the persistence of the chair. A temporal part of the chair, then, is the relativization of the chair to a particular time $t$, or to a time-interval $<t_1, t_2>$ where the chair at time $t$, or the time-interval $<t_1, t_2>$ are numerically distinct entities from both the chair and other temporal parts of the chair.

The perdurance view, then, can be summarised as saying that, for any ordinary persisting object $A$ it is true that $A$ is an aggregate composed of the series of temporal parts $<A_1, A_2, \ldots, A_n>$ where each temporal part of $A$ is distinct from $A$ and each other temporal part of $A$.

One reason for thinking the perdurance view to be correct is the fact that the persisting aggregate object is not something which we can perceive and have direct acquaintance of, since it exists at different times. What we in fact can be in direct perceptual contact with is always a temporal part of an object. And in

\(^{64}\) Heller, 'Temporal parts of four dimensional objects ', p. 325.

\(^{65}\) Ibid. p. 323.
the same way that we are able to refer to a whole complex object at time \( t \) by referring to a part of the object, so it seems reasonable to suppose that we are able to refer to the whole temporally extended object through referring to some of its temporal parts.\(^6\) It is in virtue of acquaintance with temporal parts of objects that we seem to be able to refer to and have knowledge of objects that are extended in time. This means, then, that there exists no difference between the ambiguity of part-whole relations of objects existing at a time and part-whole relations of objects existing over time. Both cases involve a certain amount of ambiguity concerning what object one has referred to in an act of referring. In the case of the spatial part-whole relation the ambiguity resides in whether the object referred to is the whole object or a certain spatial part of it. In the case of the temporal part-whole relation the ambiguity resides in whether the referring object is the in time extended object or only a temporal part of it.

A consequence of the ambiguity of reference is that one must keep track of what kind of entity one is intending to refer to in attributions of identity, since there are always two different types of entities to which identity can be attributed. Identity can be attributed to each individual temporal part of an aggregate and to the aggregate object existing in the four-dimensional space-time continuum. But, one cannot consistently attribute identity of the object as a three-dimensional object. As a three-dimensional object, the object is nothing but a succession of numerically distinct temporal parts. Three-dimensionally there is a plurality of objects and attributing sameness to that plurality is to attribute 'gen-identity' between the succession of temporal parts, i.e., what it is that makes two distinct temporal parts be parts of one and the same four-dimensional object.

Just as ordinary three-dimensional physical objects can be relativized to a time, and therefore have temporal parts, so also a person can have temporal parts. Persons persist over time in the same way as ordinary physical objects do and in this respect they are also four-dimensional objects consisting of three-dimensional temporal parts. According to this view, you and I, and all other existing persons are processes, or events in a space-time universe. In other words all persisting persons, i.e. the individuals that have a history, are four-dimensional objects, or processes in a space-time universe, and the temporal parts of persons are three-dimensional manifestations of those objects.\(^7\)

---

\(^6\) Heller, 'Temporal parts of four dimensional objects', pp. 325-326.

\(^7\) This is a simplification. Some philosophers are hard to classify. Hume, for instance, understands the person as having temporal parts, as well as having spatial parts, but he does not seem to presuppose that persons are four-dimensional objects. On the other hand, Hume does take the person as a fiction. Perhaps the best way to understand Hume is to read him as...
The temporal parts of a person are usually called *person-stages*, or *total-temporal-states* as opposite to time-related parts of objects other than persons. A person-stage or a total-temporal-state, then, can be described as the characterisation of an existing person P at the particular time t and a persisting person is "is an aggregate of person-stages". As in the case with ordinary physical objects different person-stages are numerically distinct from each other and the persisting person.

As in the case with ordinary physical objects, a person-stage could be understood as either an instantaneous time-slice of a person or as an extended time-sequence of a person where the temporal part of the person is identified with the properties held by a person relative to a time or time-interval, or a physical object filling a certain space-time interval. A *time-slice* of a person is then the momentarily existing three-dimensional instantiation of the four-dimensional persisting person. It is the person P as he is at the time t. A *person-sequence* is then the enduring temporal part of a larger persisting person. It is the person P as he is during the times t₁ to t₂. Both understandings of the person-stages entail that a persisting person P consists in the continuation of its person-stages. A person persists as long as the person-stages of which it is composed continues to exist.

### 3.8 Criticism of the perduring understanding of persons

Let me now criticise the perdurance view. The criticism against the perdurance view is not conclusive, since most of the objections levelled against it can be incorporated and made compatible with it. Though the objections cannot be understood as providing a conclusive argument against the perdurance view, they still provide us with a *prima facie* reason for thinking that the conceptual scheme of enduring objects, including ourselves, are primitive in relation to the conceptual scheme of perduring objects.

---

68 Lewis, 'Survival and identity', p. 22.
70 The problem of personal identity over time is then viewed as the problem of specifying the necessary and/or sufficient relation (or relations) that must obtain between two distinct person-stages or total-temporary-states, P ' at t' and P " at t", for those to be person-stages or total-temporary-states of the same person, P.
Firstly, the perduring view seems to entail the possibility of a complete referential indeterminacy, since we are able to simultaneously talk of, and make reference to both a persisting object and some temporal parts of that same object. For instance, we have a person, Smith, and let us assume that Smith exists between 1959 and 1994, i.e. he dies at the age of thirty-five. During his existence Smith does a lot of things: as a youngster he goes to school, plays games, later he falls in love, gets a job, etc. At the same time that we have Smith-the-persisting-person existing from 1959 until 1994, we also seem to have two much more limited person-stages, Smith-at-t', where he plays games, and Smith-at-t'', where he gets himself a job.

If Smith-at-t' and Smith had been one and the same numerical entity then they, according to the congruence-condition, should have had the same properties, since Smith-at-t' and Smith in that case would be qualitatively indiscernible. But, obviously Smith-at-t' and Smith are discernible, since Smith has a lot of properties that the temporal part Smith-at-t' actually lacks. This implies that the referent of the expression 'Smith-at-t'' and the referent of 'Smith' are not the same entity. The same is true of the other person-stage, Smith-at-t''. But, this means that in every referring expression we can make reference to more than one entity at every specifiable time. For instance, in statements such as ‘That person over there is annoying me’ or ‘Jones visited London yesterday’ it is indeterminate whether we are referring to the continuant person, or whether we are in fact talking about a particular person-stage of that continuant person. This indeterminacy of what we are in fact referring to with our sentences would be equally indeterminate when we are making first-personal references, like ‘I am angry!’ or ‘I ate three apples yesterday’. It is not strictly clear whether I am talking about myself as a persisting object, or whether I am talking about certain stages of myself. I do not know the intuitions of other people, but there certainly seems to be something wrong with a theory which entails the possibility of an internal or intra-subjective referential indeterminacy in every referential situation using ordinary language.

Another problem with the perdurance view is overpopulation. If we accept that persons are objects that persist in virtue of having temporal parts, then, several distinct entities will exist at the same place and time whenever there exists a person-stage which is the person-stage of a person. On the one hand, there will be the person-stage existing at t, and on the other hand, there will be the persisting person with a history terminating with the person-stage at t. But, accepting this is difficult, to say the least.
Another problem, also due to the fact of dual reference, is counterfactuals. Consider such a simple statement as; 'I could have died when I was ten'. Who is this statement about? Well, obviously, it seems to be about myself. I am imagining a situation that could have occurred, and in that situation I am imagining that I, who am writing this sentence now, would be dead. It even seems to me now, that I can see myself falling from a cliff, see the fear in my face, etc. But, this is wrong according to the perduring view. It is not the case that I can imagine myself being dead. This is not due to any special problem about imagining the non-existing. It is because that person simply would not be me. Let us assume that the continuant person that I am is constituted out of the person-stages $S_1$, $S_2$,....,$S_n$. Let us call this person 'I'. The continuant person, supposed to be me, who died at the age of ten would be constituted out of the person-stages $S_1$, $S_2$,....,$S_n$. Let us call this person 'RoM'. Obviously, I consist of more temporal parts than RoM, since I have been existing much longer than RoM. Since I consist of some parts which RoM lacks, I will also have some properties that RoM does not have. I cannot therefore be the same process, or persisting person as RoM. The same fact also implies that I cannot really be thinking of myself when I have future-oriented thoughts involving expressions such as ‘I’ and ‘Roger Melin’. But, it seems pretty strange to accept a view of myself which implies that I simply cannot think of, or at least not consistently think of, myself in the past or in the future.71

The last objection against the perdurance view is that it is odd as an account of change in objects. As an account of the possibility of change in objects the proponent of the perdurance view maintains that change is nothing but the occurrence of different, but qualitative similar temporal parts of an aggregate which, looked upon from an eternal point of view, have not changed. So, when we are asserting that an object $O$ has changed, what we have done really is to assert that a particular temporal part $t$ of the aggregate object $O$ has been succeeded of a completely different particular temporal part $t'$ of that very same aggregate object $O$. But, to claim that ordinary three-dimensional objects simply do not change is contrary to all experience. What we experience is that a single

71 This criticism could, of course, be countered if one accepts a “Counterpart theory”. (See, for instance, David Lewis, ‘Counterpart Theory and Quantified Modal Logic’ in Philosophical Papers: Vol. I, (New York: Oxford University Press, 1983). But, personally I do not see any convincing reason why one should accept such an account of counterfactual statements. It does, for instance, make more sense, I think, if counterfactual statements about how my own possible life could have been are understood as literally being about the individual that I am, and not about a different life.
object is passing through changes, not that a non-observable four-dimensional object remains through the change.

3.9 The persistence of persons: the Endurance view

As I stated earlier, the idea of an enduring object is that of an object which is "wholly present" at every single moment of its' existence. Ordinary objects are three-dimensional entities which remain in existence even though they acquire new properties from time to time. In claiming this, an advocate of the endurance view upholds the idea that there is a sharp distinction between objects and the processes and events of objects. While events and processes do extend over time by having temporal parts that are numerically distinct from each other, our ordinary objects, on the other hand, can only have spatial parts. Ordinary objects, though existing at different times, persist through time in virtue of being numerically the same object at every time that object exists.

In this sense, since it is one and the same object which exists at the different times $t_1$ and $t_2$, it is claimed that one should literally understand statements about possible courses of events as referring to one, and only one, single object. For instance, in claiming that the computer that I am now writing this sentence on could have been destroyed three months ago, what I am claiming is that this very computer that I have in front of me could have been destroyed three months ago.

Applied to persons, the endurance view consists in claiming that when a person $P$ persists between $t$ and $t'$ it is numerically one and the same person that is present at $t$ and at $t'$. Persons are three-dimensional objects that can acquire different properties at different times of their existence. When I express statements such as 'I went fishing when I was five years old' or 'I might have died four years ago' what I am referring to is not certain temporal parts which are distinct from me, but to myself as I was when I was five years old and what could have been the case four years ago. That is, what I am referring to in these statements is literally speaking, that it was I, the very same object that I am today, that could have passed through these alternative courses of life.

The problem of the enduring object having different, and even incompatible properties at different times can be solved if we accept that certain properties of objects are accidental, while others are essential, and by also taking tenses seriously.
Accepting the distinction between accidental properties and essential properties makes my version of the endurance view Aristotelian in spirit in that it takes persons to be "primary substances" or continuants who persist through time and change. According to this view, there is a logical, or conceptual distinction between, on the one hand, primary substances, and, on the other hand, modes of such primary substances. A primary substance is an ordinary object, such as an individual dog or a person, while a mode is either a process, an event or a property of such an object. Two important points about the primary substances are that they are "the ultimate subjects of predication" and that they have a certain distinctive unity independent of our classificatory scheme.

Though it is true that a property such as walk can function as the subject of predication, as in 'The walk was quick', it is the case that an individual instance of walking cannot be thought of as existing in isolation from some walking thing, i.e. a primary substance which is taking a walk. In this sense, individuating a mode is always dependent upon a prior individuation of a primary substance.

As to the question of unity, modes, contrary to primary substances, do not have a natural unity or boundary, but are always unified according to our decisions. For instance, consider the action of my walking down to college. In the walk itself there is nothing which absolutely decides when and where the walk begins and ends. Is my walking towards the door of Exeter House a part of my walking down to college or not? What if I stop on Iffley Road to get tobacco and cigarette-paper. Is that to be counted as a part of the walk? Maybe it is even possible, as the last factor indicates, that one and the same walk can disappear for a certain time and then restart again. What this shows is that, in the walk itself, there is nothing which determines what to count as a part of the walk, and that it is we, by using our classificatory capacities, who determine when and where a certain walk begins and ends.

While this is true of modes it is not true of substances. Individual substances have a unity which springs from the objects themselves. They have, as Michael Ayers claims, a real unity determined by reality itself. Consider for instance a

72 Talking about 'primary substances' implies that I, more or less, accept the Aristotelian account of substances as it occurs in Aristotle, * Categoriae, in Categorieae and De Interpretatione*, Translated by E. M. Edghill, (Oxford: Oxford University Press, 1928), Ch.5. In this sense, a substance is an individual object instantiating sort-determining characteristics partaking in the (spatio-)temporal order.


horse. The horse is a member of a biological species. We classify the horse as belonging to its particular species on the ground that it has a certain kind of internal and morphological structure based upon its particular origin (that it is the offspring of two horses).\textsuperscript{75} In this sense, the horse's gross\textsuperscript{76} morphological structure is to be explained by the fact that it is caused by the internal, or genetic structure of the object, and the internal structure is to be explained as caused by its causal and evolutionary history. Since the internal and morphological structure of the individual horse depends upon its ancestors, it is a horse, not in virtue of having certain features, but by having the right causal and evolutionary history. In this sense, an individual horse is the product of evolution, and it has its fundamental properties in virtue of the evolutionary process. If, for instance, evolution had taken a slightly different route than it actually did we may not have had any horses at all. The unity that individual horses have is an effect of evolution, and can be said to exist in the individual horse itself.

At this stage I will take it that persons, whatever else they are, are suitable candidates for being the ultimate subjects of predication, and that they do have a certain distinctive unity in themselves which would explain the development and growth of persons in a similar way to that of the individual horse. It is, for instance, appropriate to predicate certain mental states such as a particular pain or experience of an individual person. And even though we can make predications of mental states, for instance 'The pain is terrible', this we can do because we have a prior individuation of the mental state as being the state of some individual person. Similarly with the case of \textit{walk}. It is commonly the case that persons are taking walks, and that our ability to individuate a single walk is due to our ability to individuate a person who is walking. Whether it is the same walk a person is taking after he has stopped for an hour at the coffee-shop on High Street, or whether he is taking a new walk is very much a matter for discussion and convention. But, whether it is the same person who is walking after stopping at the coffee-shop is not subject to conventional decision. Whether it is the same person or not does not, as it seems, depend upon whether we \textit{want to call} it the same person, but upon whether it \textit{really} is the same person who is going for a walk. In this regard we have at least some indication that persons are primary substances.

\textsuperscript{75} Ayers, Locke: Volume II Ontology, p. 224.

\textsuperscript{76} I say 'gross' here, since it is possible that a horse's morphological structure can depend upon things beside the genetical structure and evolutionary history. For instance, if we have to cut off one leg of a horse we change the morphological structure of the horse, but this new morphological structure does not depend upon the internal structure or evolutionary history of the horse.
Secondly, let us assume that it is possible to construct a sortal term covering every possible property of an object. For instance, there is the sortal term ‘red-haired’ covering the property of having red hair and the sortal term ‘green table’ covering the property of being a table coloured green. These sortal terms are phased sortal terms, since they pick out a property of an object which the object can loose without ceasing to exist. But, this implies that one simply cannot treat the phased sortals in an atemporal manner. Every sentences involving phased sortals are a form of implicit tense expression. In this sense, every phased sortal can be related to some specific time at which it is truly applicable of the object. Statements involving the attribution of contradictory phased sortal terms should always be understood as entailing an implicit tensed component expressible by ‘was’ or ‘will become’ or something similar. In such a statement it is one single object which has properties at different times and it is only if one overlooks the entailed tensed character of the statement that these properties can be said to be incompatible.

This is in sharp contrast to the perdurance view. Suppose that we have a situation in which a certain object $a$ is attributed both the property $p$ and the property not-$p$. An advocate of the perdurance view would have to claim that this is impossible, and conclude that there have to be more than one object involved. One object having the property $p$, and another object having the property not-$p$. Contrary to that, and more consistent with common sense and ordinary language, I claim that it is impossible for an object to have the properties $p$ and not-$p$ at the very same time. But, there is nothing wrong in the idea that one and the same object can have $p$ and not-$p$ at different times of its existence. This means, then, that whenever we are faced with sentences attributing incompatible properties of an object, we should closely analyse the sentence in search of some form of temporal misuse. Being sensitive to the tensed character of phased sortal terms, there simply is no need to postulate objects for which we have no empirical evidence at all. It is only by overlooking the tensed character of certain expressions that make it seem as if an ordinary three-dimensional object, i.e. a primary substance, is an impossible object.

---

3.10 The ontologically priority of the Endurance view

I have already argued that something must remain numerically the same over time if a denial of numerical sameness of objects over time is to make any sense at all.\textsuperscript{78} What I claimed was that the denial of numerical identity of objects over time must presuppose a conceptual scheme where certain fixed entities continue to persist. What I further want to show is that the conceptual scheme which we are using also requires that ordinary objects, particularly persons, remain numerically the same over time.

The reason for claiming this is that when advocates of the perdurance view are to define what it is to be a momentary thing they make use of ordinary terms such as 'horse', 'tree' and 'person'. For instance, a momentary thing such as this-person-at-4th-of-July-1997 is defined as the set of space-time co-ordinates \(<o_1, o_2, o_3, t>\) (where 'o' stands for occupying a space and 't' stands for the time), such that there is a point occupied at the 4th of July 1997 by a certain portion of matter which at the 4th of July 1997 constitutes this person. The four-dimensional person is then understood as being the series of these momentary things.

But, as Wiggins has pointed out, it is an essential feature of the process of defining a momentary thing that the three-dimensional language comes in.\textsuperscript{79} In the definitions the expression 'this person' refers to the three-dimensional person, contrary to what might be thought. It cannot refer to the four-dimensional object since that object is supposed to be the sum of all those momentary stages. That is, in defining what a person-stage, or a time-slice of a person is, it is necessary that we make use of the three-dimensional notion of a person. In this sense, stages of persons are abstractions of three-dimensional persons, or continuant persons, and we need to master the three-dimensional understanding of persons before we can even start constructing a four-dimensional view of ourselves. As Wiggins claims, "The definitional priority of the continuant language, in which the construction of four-dimensional counterparts of three-dimensional continuants is founded, is instantly and unreservedly conceded by anyone who candidly adopts it".\textsuperscript{80}

Secondly, given that it is possible for an advocate of the perdurance view to define what it is to be a momentary thing, the perdurance view is still in trouble with predication. If the perdurance view is to be a possible candidate for really

\textsuperscript{78} See Chapter 2, Section 2.4.

\textsuperscript{79} Wiggins, \textit{Sameness and Substance}, p 195.

\textsuperscript{80} \textit{Ibid}. p 195.
competing with the endurance view, it must be possible to translate all talk about ordinary three-dimensional objects into talk about four-dimensional objects. That is, it must be possible to find a formulation in which what is truly applicable to an ordinary continuant comes out as true also of the four-dimensional object. But, as Wiggins claims, this means that for every predicate $F$ in our ordinary language, there must exist a true biconditional $A \leftrightarrow B$ such that (1) $A$ was constructed with the help of $F$ and without the help of expressions not belonging to the continuant language itself; and (2) $B$ was constructed from expressions belonging to the four-dimensional language and from no expressions belonging only to the portion of the continuant language that was to be reduced or interpreted.\footnote{Wiggins, *Sameness and Substance*, p 196.}

Take, as an example, the predicate ‘is eating’. We all apply this predicate to ourselves and to other persons (as well as to certain animals like dogs and cats). First of all, it cannot be the case that a momentary instance of a thing can eat. That is, ‘is eating’ is not a predicate which is applicable to a stage of a person, since eating is a process which necessarily takes some time. Furthermore, what is required if the perdurance view is to be independent of the endurance view is that it should be possible to find a predicate, or to construct a new predicate in our ordinary language which corresponds to a momentary instance of eating, and then find a predicate in the four-dimensional language which corresponds to the momentary instance of eating, but which at the same time is not defined through ‘is eating’. But, since we have seen already that momentary instances of things cannot be defined without using the language of three-dimensional objects we will not find any instances of such bi-conditionals. What we can accomplish, then, is at best a definition of a four-dimensional object which relies upon three-dimensional objects for its intelligibility. In this sense then, the endurance view of ordinary objects is conceptually-cum-ontologically prior to the perdurance view of objects.

A last argument in favour of the endurance view in relation to the perdurance view can be extracted from remarks due to Saul Kripke.\footnote{I am mostly relying on a report of Kripke’s argument by Sten Lindström and Włodzimierz Rabinowicz as it appears in ‘Kripke in Uppsala’ in *Filosofisk Tidsskrift*, No 3, 1986, pp. 35-40. The text is a report of Kripke’s Hägerström Lectures in Uppsala 1985. The argument I am presenting can be found in Saul Kripke, ‘Identity over time’ Unpublished manuscript.} Suppose the world to consist of a flat homogenous circular disk. Let us further imagine that the disk is single-coloured, say white. Given that such a disk exists, we can intuitively distinguish between a case in which the disk is rotating and a case in which the disk
is still. These two cases are distinguishable, since they intuitively describe two different facts.

But, this distinction is not possible to uphold if one adopts the perdurance view of objects. If the perdurance view is correct, then the disk is a four-dimensional object, i.e. built up from momentary stages that all qualify as being stages of one and the same object, the disk, in virtue of the existence of a qualitative resemblance between closely neighbouring stages. But, since we have assumed that the disk is one-coloured and homogenous, each stage of the disk must be qualitatively indistinguishable from all the other stages of the disk regardless of whether the disk rotates or not. This implies that the perdurance view cannot differentiate between the situation where the one-coloured homogenous circular disk remains exactly similar through its rotating movement, and the situation where the disk remains exactly similar and is at rest.

Since the endurance view can make a distinction between what is intuitively judged as two different situations, while the perdurance view cannot, this implies two things. Firstly, since we actually do have the intuitive distinction between the rotating disk and the unmoving disk, the endurance view must be understood as primitive in relation to the perdurance view. Secondly, the endurance view has more expressive power than the perdurance view, since it provides us with a way of distinguishing between situations intuitively thought of as being distinct. In this sense, though the perdurance view can translate many sentences of the three-dimensional language into a four-dimensional language, the perdurance view cannot express everything expressible in a three-dimensional language. There are common-sense facts, intuitively judged as obtaining, which the perdurance view cannot express. But, since the perdurance view cannot express what is intuitively judged as obtaining, it cannot be an adequate account of how objects persist. Hence, the endurance view is the basic view of our conceptual scheme.
4. CRITERIANISM AND NON-CRITERIANISM

4.1 Introduction

In this chapter I will be looking at what I take to be the two opposing traditions of personal identity: Criterianism and Non-Criterianism. Criterianism claims that the identity over time for objects like us consists in psychological and/or some form of physical continuity and has been endorsed by a wide variety of philosophers, such as John Locke, David Hume, Derek Parfit, John Perry, Peter Unger and Michael Ayers. Non-Criterianism, on the other hand, upholds the view that the psychological and/or physical continuity which accompanies our identity is, at best, epistemological evidence for the fact of our identity over time, but that the fact of identity over time for us as being persons is primitive in relation to these empirical relations of continuity. Non-Criterianism has been endorsed by philosophers such as René Descartes, Thomas Reid, Joseph Butler, and in more recent times, by Roderick M. Chisholm, Richard Swinburne, Geoffrey Madell and E. J. Lowe. But, from what we said in the introduction,

---


both Criterianism and Non-Criterianism are so far without any real content. In presenting different versions of Criterianism and Non-Criterianism I will determine essential similarities and contingent dissimilarities of particular theories of the two opposing traditions.

4.2 Criterianism

The Criterianism of interest for this inquiry can be divided into two different groups. On the one hand we have philosophers who are Criterianists with regard to the predicate 'personal identity' and on the other hand there are Criterianists with regard to the predicate 'our identity'. The general idea behind this distinction is that certain philosophers think that it is wrong to presuppose that we are essentially persons. Personal identity, they think, is an intricate and difficult matter which does not necessarily have an determinate answer. According to these philosophers it is easier to determine what is essentially involved in cases of identity over time for objects like us if we disregard the concept person from the discourse and put all our focus upon ourselves.

As I claimed in the introduction, personal identity, or our own identity over time, is something which we experience as a phenomenological fact. That I am identical with a certain individual of the past is immediately given to me in my experience of myself as a persisting object. I do not, as it seems, infer the persistence of myself from knowledge of certain other and more basic facts. It seems, then, that personal identity is experienced as a primitive fact.

This primitiveness of personal identity is exactly what Criterianism denies. Advocates of Criterianism do not necessarily deny that personal identity is a fact, but they necessarily deny that personal identity is primitive. The reason for this is that advocates of Criterianism think it is possible to specify a criterion of identity for persons or objects of the same sort as you and I. By this, Criterianism entails that (i) it is possible to specify a strong criterion of identity for persons in that the essential conditions for the fact of personal identity is specified in terms which only mention facts other than personal identity, and (ii) these facts are empirical facts with an epistemological privileged status in relation to statements about the experienced fact of personal identity. The specificative procedure advanced by Criterianism can take different forms, but the idea is that we can specify a non-empty set of conditions $C_1, \ldots, C_n$ such that, if and only if the conditions $C_1, \ldots, C_n$ obtain, do we have identity over time for objects like ourselves.
Let me now further elucidate Criterianism. In doing this it will be helpful if I first introduce a distinction between two different classes of statements of facts: the given class and the specificative class. The given class consists of all those statements of facts which we are prepared to accept without reflection. It is those statements of fact with which we are presented prior to any analysis. The specificative class, on the other hand, consists of those statements of facts which we accept as the conceptual or ontological ground for the statements in the given class. It follows that the specificative class is formed by the statements with which we end our analysis. That is, the specificative class consists of statements of necessary and sufficient conditions for the existence of facts expressed by statements in the given class.

This specification should be conducted in an empiricist framework. The reason for this is that, in having a criterion of identity for objects of a certain kind K, we are supposed to ground our talk about identity over time of K:s in something more basic. But, the determination of statements of the given class cannot be done if the specificative class does not refer to observable facts of the world. The fact of specifiability implies that there exists a "method of verification" in which it is possible to establish that the specified facts obtain. That is, not only is it enough to claim that it is possible to specify any facts whatsoever, it is also required that the specification should rely upon conditions whose satisfaction can be determined. It is, furthermore, in virtue of these conditions, expressed in the statements of the specificative class, that the statements of the given class get their significance. In this sense, the statements of the specificative class should consist of expressions referring to empirical facts. It is not until we have matched the facts of the given class with empirical conditions of the specificative class that we can say that we have made a specification of the facts of the given class.

The general idea of Criterianism with regard to personal identity or our identity consists in upholding the following four different theses:

(i) The specifiability thesis, i.e. that it is possible to specify necessary and sufficient conditions for personal identity or our identity,

(ii) The derivativeness thesis, i.e. that the fact of personal or our identity can be semantically-cum-metaphysically understood if and only if it is related to some more basic facts,

This distinction goes back to Michael Dummett's defence of anti-realism, but my use of the distinction in the context of Criterianism does not carry with it any particular position concerning anti-realism or realism. I use it only as a mean of illuminating the Criterianist position. See Michael Dummett, 'Realism', Synthese, vol. 52, 1982.
(iii) The complexity thesis, i.e. that the fact of personal or our identity depends upon some more basic facts, and

(iv) The essentiality thesis, i.e. that the specified necessary and sufficient conditions for personal or our identity should capture what is essentially involved in cases of identity over time for objects like us.

4.3 The specifiability thesis

Let me illustrate the idea of the specifiability thesis. Suppose that we know that a ship $S$ exists at $t_1$ and that a ship $S'$ exists at $t_2$, and that we, as we are used to say, know that the ship $S'$ is the same ship as, or is identical with, the ship $S$. Now, the identity, or sameness of $S'$ with $S$ is a fact which, in this case, is expressed in our given class. The question we can ask ourselves, then, is whether there are any observable or empirical conditions that we could specify which would make it true that if and only if these conditions are satisfied would we have a case of ship-identity between $S'$ and $S$. In a general way, this question will be answered if we can specify some more basic facts such that $S'$ at $t_2$ is the same ship as $S$ at $t_1$ if and only if these basic facts take place during $t_1$ and $t_2$.

Given that this is a oversimplified version, we would be inclined to say that $S'$ is identical with $S$ if and only if $S'$ consists of matter which is spatio-temporally connected with the matter of $S$, and there is no other ship $S^*$ existing at $t_2$ which is also spatio-temporally connected with the matter of $S$. In this sense, the condition that we specify, i.e. spatio-temporal continuity, is the condition which is necessary and sufficient for $S'$ to be identical with $S$.

---

70 'Spatio-temporal continuity' does not here mean that $S'$ must consist of the same matter as $S$, since that would imply that every repair of the ship would count as the destruction of the ship. What it means is that a replacement of the matter of the ship is possible while still retaining the ordinary ship if the replacement of matter is gradual.

71 It is a necessary condition, since it seems very odd to accept that a ship consisting of the matter $M$ at an instant later could consist of the matter $M'$. If we have a ship $S$, consisting of matter $M$, and an instant later has a ship $S'$ consisting of $M'$, then $S'$ is not the same ship as $S$. And, as ordinarily understood, and ignoring some possible thought experiments, spatio-temporal continuity is a sufficient condition for ship-identity, since having the same, or more or less the same matter is enough to count as being the same ship. For instance, if a sailor leaves the harbour in the morning and during the time he is out on the sea has to replace certain smaller parts of his ship, we do not conclude when he arrives in the harbour the next day that he is sailing another ship than the one he left the harbour with.
We have, then, through specifying spatio-temporal continuity of the matter of the ship as being necessary and sufficient for the persistence of ships, specified the criterion of identity for ships through relating the given class statements about ship-identity to the specificative class statements about spatio-temporal continuity.

In the same way that the identity over time of ships can be specified by relating it to a certain form of empirical continuity, it is thus assumed that personal identity or our identity can be so related. In this sense, there is in principle, no difference between our identity over time and identity over time of other kinds of objects, such as ships, cars and trees. According to Criterianism, in both cases there is a specification of statements of the given class into those of the specificative class. Though, the Criterianist continues, the specified conditions for our identity over time may be other than the conditions needed for a ship's identity, since individuals like us do have features which ordinary material objects lack.

Some Criterianists have taken the fact that persons have a mental life as the overall feature of persons, and concluded that the important condition for persons is psychological continuity, while other Criterianists, though not denying that persons have a mental life, still think that the important conditions for personal identity are some kind of spatio-temporal continuity. I call the first Psychological Criterianism and the latter Physical Criterianism. A third possible Criterianist account is the account which is Criterianist with regard to the predicate 'our identity'. This view I call Animalism. According to this view, while it is true that persons have a mental life, this is not in any sense essential for being the individuals that persons are. For instance, I am fundamentally an animal and the important condition for remaining myself over time is that of remaining the same animal.\(^{72}\)

\(^{72}\)It is important to notice that more recent versions of Psychological Criterianism also claim that physical continuity is of importance to personal identity over time and that Animalism could uphold a view in which psychological continuity is, at least, a sufficient condition for personal identity. Recent Psychological accounts claim that the continuity of physical parts, either the whole body, the brain or the central nervous system, is a necessary condition for the existence of psychological continuity. In this sense, psychological continuity incorporates a necessary physical condition. This applies also to Animalism. They might uphold a view which claims that personal identity consists in psychological continuity which is grounded in the same animal. The reason for me to separate these as being three different versions of Criterianism is that, even though it is true that recent versions of Psychological Criterianism as a matter of fact involves physical continuity as a necessary condition and that Animalism could entail a criterion of identity for persons, this is not something which a Psychological Criterianist or an Animalist has to do. It is possible for a Criterianist to maintain that the only condition necessary for personal identity is psychological continuity of some sort. As a matter of fact, it is even the case that some Psychological Criterianists of the past, for instance John Locke and David Hume, seem to have had such a pure psychological continuity view of personal identity. A reason for many Animalists to hold their view is a scepticism concerning
Though Criterianists may differ regarding the exact specification of necessary and sufficient conditions, they all agree in claiming that the only difference between the case of the ship and the case of persons is that persons, contrary to ships, trees and cars have a psychological life.

4.4 The complexity thesis

Consider once again the similarity between our understanding of our own identity over time and our ordinary understanding of ship-identity over time.

In claiming that the ship S' is spatio-temporally continuous with the ship S, and that it is this fact of spatio-temporal continuity between S' and S, that makes S' identical with S, we have in a sense reduced the fact of the identity over time of the ship to the spatio-temporal continuity of the matter constituting the ship. Though the ship is, as we say, identical over time, this identity over time is nothing else than the spatio-temporal continuity which the ship displays. The fact that the ship S' is identical with S is nothing over and above the fact that S' is spatio-temporally continuous with S. We can express this by saying that talk about ship-identity implies talk about spatio-temporal continuity. Whenever we describe the fact of ship-identity over time we also describe the fact of spatio-temporal continuity, either because we describe the same fact at different levels of description or because certain descriptions of facts imply the existence of other and more basic facts. In the first case, describing the fact of ship identity in statements of the given class is just another way of describing the same fact expressed by the statements of the specificative class. In the second case, describing the fact of ship identity in terms of statements of the given class implies the truth of the statements about spatio-temporal continuity expressed by the statements in the specificative class.

In the same way that we can specify the identity over time of a ship in terms of some more basic facts, a Criterianist argues, describing the persistence of a person implies that some more basic events or facts obtain. In this sense, personal identity or our identity is nothing over and above the fact of these more basic events specified by the statements of the specificative class. By accepting the possibility of establishing a general criterion of identity for persons. Because of this scepticism they think we ought to change the focus from viewing ourselves as essentially persons to viewing ourselves as essentially animals or human animals.

73 Parfit, Reasons and Persons, p 213.

74 Ibid. pp. 210-212.
this idea about specifiability into more basic facts, a Criterianist also accepts the complexity thesis. That is, the statements of the specificative class refer to the basic facts, and what is talked about in the statements of the given class just consists in the holding of some more particular facts. For instance, though it is true that we can talk about personal identity over time, and that the statements of the given class imply statements of the specificative class, or that these different classes of statements express one and the same fact, statements about identity over time are not the basic level of description of the fact of a person's persistence. Identity over time of a person is, in this sense, a complex fact.

The idea behind a fact being complex goes back to the Lockean distinction between simple and complex ideas. According to Locke, an idea is simple if it contains "one uniform appearance, or conception of the mind", while an idea is complex if it consists of several simple ideas united together. Drawing on this distinction, what is simple, according to the Criterianist, are facts which can exist in and by themselves, while what is complex are facts which are built up by simple facts. And, the Criterianist continues, the basic, or simple facts in cases of our identity over time are mental and/or physical states or events specifiable in the specificative class, and not the identity of the person which we express in our statements of the given class. In this sense, the specification of the fact expressed by statements of the given class in terms of statements of the specificative class will consist of the specification, or clarification of something complex in terms of something simple, or more basic.

That Criterianism entails the complexity thesis rests upon the fact that Criterianism emphasises that the specified conditions of the specificative class jointly suffice as the criterion of identity over time. Consider, for instance, once again our very simplified version of the identity of ships. In asking what the identity of ships consists in we do not want our answer just to be of the form; "S₂ at t₂ is the same ship as S₁ at t₁ if and only if S₂ is the same ship as S₁". Though being absolutely true, this answer cannot be accepted as a criterion of identity for ships, since it is not informative. It is a reasonable requirement of a criterion of identity that a criterion of identity for objects of a certain kind K must involve a commitment to non-circularity and informativeness. A criterion cannot, if it is to

---

be of any use as a criterion, already presuppose, or make reference to the kind of thing of which it is a criterion. Thus, stating a criterion for objects of the kind K in terms of K-identity cannot be a genuine criterion for objects of the kind K, since the purported criterion does not give us any more knowledge than what we already had. In our particular case, we already knew that $S_2$ was identical with $S_1$. What we also wanted to know was what this identity between $S_2$ and $S_1$ consisted in. What we need to do if we want to find a criterion of identity over time for ships, is to specify some more basic, or more lower level facts that obtain whenever the identity of a ship obtains.

What is true of the identity over time for ships is also true of personal identity. The specified criterion which is specified should not presuppose, or make reference to the kind of thing of which it is a criterion. In this sense, the specification of necessary and sufficient conditions for what our identity over time consists in should be \textit{weakly impersonal}. That is, it must be possible to describe the facts constituting the necessary and sufficient conditions without presupposing the identity of the individual under consideration.\footnote{Parfit, \textit{Reasons and Persons}, p. 210.}

\section*{4.5 The derivativeness thesis}

Related to the fact that Criterianists think that the fact of our identity over time is nothing \textit{over and above} the specified empirical continuities is the thesis that the fact of our identity is, in some sense, derivative. We can have knowledge of the statements of the given class, i.e. statements expressing facts of personal identity over time, only because these facts are dependent upon the conditions specified in the specificative class. It is in virtue of an implicit grasp of the scheme of specification that we can actually understand the statements of the given class.\footnote{Dummett, \textit{Realism}, p. 66.} What Criterianism claims in accepting the specifiability thesis, is that statements involving ‘personal identity over time’ implicitly mean, or imply the holding of certain more particular facts, and that it is in virtue of the implicit grasp of the scheme of specification proposed by the Criterianist that we can actually understand the statements about personal identity over time.\footnote{\textit{Ibid.} p. 66.} In this sense, Criterianism maintains that our identity over time as persons relies, for its intelligibility, upon the specified conditions of the specificative class and this
fact is the foundation for our talk about personal identity over time as a fact of the world.

A Criterianist must accept the derivativeness thesis on the ground that the criterion given by a Criterianist must be informative. Since it is necessary that Criterianists uphold the two claims (i) that our identity over time can be specified in terms of a criterion for personal identity or identity for objects of our kind, and (ii) that the criterion, if it is to be informative, has to be formulated in some more basic facts, then Criterianists also uphold the idea that the fact of personal identity or our identity is derivatively dependent upon these more basic. That is, according to Criterianism, the primitive facts are those that are specified in the statements of the specificative class, while the statements of the given class which express the fact of personal identity over time are conceptually and ontologically secondary and rest upon the basic facts of the specificative class.

4.6 The essentiality thesis

A further necessary claim of Criterianism is the essentiality thesis. By the essentiality thesis a Criterianist accepts that the specified conditions which are said to give the criterion for our identity should not only specify any facts whatsoever that always accompany our identity over time, but that it should capture what essentially is involved in cases of identity over time for objects of our sort.

The essentiality thesis is a direct consequence of the fact that a criterion for the identity over time of an object of the kind K should specify semantic-cum-metaphysical conditions for what it is to be an object of the kind K. The specified necessary and sufficient conditions should, in this sense, give a partial definition of what it is to exist over time as an object of our sort, and not only specify some conditions by which we come to know that such identity over time obtains.

So far, then, we have seen that Criterianism accepts that it is possible to specify our identity over time in terms of more basic necessary and sufficient conditions, and that this fact of specifiability implies that ‘personal identity’ or ‘our identity’ refers to a derivative and complex fact of the world. Let us formulate the four central claims of Criterianism in the common Criterianist thesis:

(i) personal identity, or our identity, is epistemologically and/or metaphysically derivative and complex,
(ii) there is a specifiable finite set of empirical conditions $C$ that are necessary and jointly sufficient for an object $a$ belonging to our sort to be identical with an object $b$ belonging to our sort such that

a) we can formulate a "definition" of the form: the object $a$ is identical with the object $b$ if and only if $C_1(a, b) \ldots C_n(a, b)$, and

b) the conditions $C_1, \ldots, C_n$ should not employ identity for objects of the sort.

4.7 Different versions of Criterianism

Some Criterianists claim that a question about what our identity over time consists in always has to be determinate on the ground that our identity, in contrast with the identity over time of other kind of objects, is absolute. Other Criterianists claim that personal identity can be indeterminate in truth-value. What these latter, but not the former, theories of Criterianism accept is the extra Criterianist thesis:

(v) The Indeterminacy Thesis, i.e. the fact that our identity can be indeterminate in truth value on the ground that the essential continuities and connections involved in our identity over time can hold to various degrees.

Another possible way in which Criterianists differ is whether they are entity reductionists or not. Some Criterianists claim that the existence of a certain type of object $X$ in the given class, is nothing over and above the existence of certain other type of object $Y$ in the specificative class, but that we must still talk about, and quantify over the entities $X$ in the given class. Other Criterianists explicitly claim that it is possible to eliminate the given class type of objects $X$ completely from the universe of discourse on the ground that no such things as $X$:s exist in the world, and that every statement about $X$:s could be replaced by statements about $Y$:s. What these latter, but not the former versions of Criterianism accept is the extra Criterianist thesis:

(vi) The reductionist thesis, i.e. that the fact about our identity over time can be specified in terms of an empirical relation that does not
mention the existence of individuals like us, but only the existence of experiences, bodily movements and events.\textsuperscript{83}

\textbf{4.8 The indeterminacy thesis}

Earlier I claimed that Criterianism, in accepting the four necessary claims, upholds the idea that there is, in principle, no difference between our identity over time and the identity over time of other kinds of objects. Still, when it comes to the acceptance or non-acceptance of the indeterminacy thesis this similarity between our identity and other kinds of objects can diverge. On the one hand we have philosophers like David Lewis and Harold W. Noonan who maintain that the conditions of the specificative class can diverge from the identity relation of the given class when it comes to artificially construed entities, such as bridges and clubs, while personal identity never can be indeterminate in truth-value.\textsuperscript{84} On the other hand we have philosophers like Derek Parfit who maintain that questions about personal identity might lack determinate answers on the ground that the important continuities and connections involved in personal identity over time can hold to various degrees.\textsuperscript{85} Let us call the first view \textit{Deterministic Criterianism}, and the second view \textit{Indeterministic Criterianism}.

According to Deterministic Criterianism, the identity over time of ordinary complex objects can sometimes be indeterminate in truth value at the level of the given class, since it is imaginable that the essential empirical continuity can hold to various degrees. For instance, it is easy to imagine that we could describe every possible relation on the specificative level between Germany before the second-world-war, East-Germany, West-Germany, and the Germany of 1994, and still not know the answer to the question: "Is Germany 1994 identical to Germany 1942?". So, even though we know everything there is to know about the case in terms of the relations between constitution, people, borders, etc. the question could still be indeterminate at the given class level.

But, contrary to sentences about other kinds of complex entities, in the context of personal identity over time, the Deterministic Criterianist claims, it will


always be the case that at any particular time in the future either that person exists at that time, or he does not exist at that time. The reason for this is that the essential continuities and connections underlying personal identity cannot be of lower and higher degrees. According to Deterministic Criterianism, since we cannot make sense of the idea that I will partially exist in the future and partially not-exist in the future, my identity is always an all or nothing affair, and therefore always absolute. If the necessary and sufficient conditions for our existence over time are psychologically and/or physical continuity, the Deterministic Criterianist claims that I today is as much psychological and/or physically continuous with myself in thirty years from now, as I am continuous with myself just a minute ago. Since the Deterministic Criterianist claims that our existence over time is absolute, he also claims that we cannot be left with a conceptual shadow when we are confronted with questions about identity over time for objects of our sort. What matters in survival is always both psychological and/or physical continuity between me now and some future state and "identity between myself, existing now, and myself, still existing in the future". When we have established the conditions that underlie our identity over time, i.e. the psychological and/or physical continuity of the specificative class, and established a truth-value for every sentence of that class, a sentence about our identity over time will also always be given a truth-value.

Indeterministic Criterianism, on the other hand, accepts that all kinds of complex entities, including the identity for objects like us, have a gradual existence over time. That is, the Indeterministic Criterianist claims that questions about our identity can be indeterminate in truth value, even though we know everything there is to know about the extent of essential continuity at the level of the specificative class.

That is, if we have a given class sentence A about personal identity this guarantees that there exists a specificative sentence C with a determinate truth-value. But, there is no guarantee that the truth of the statements of the specificative class A entail the truth of a given class sentence F nor that it entails the truth of the negation of F (where the given class is understood as being closed under negation). The reason for this is that we might be able to describe many cases.

---

87 Noonan, Personal Identity, p. 20
88 In claiming this, the Indeterministic Criterianist must add an extra requirement, the non-branching clause, on the specified necessary and sufficient conditions when he thinks it is a clear case of personal identity. The reason for this is that, since complex objects can persist in virtue of being more or less "identical" it needs to be ruled out that two individuals are equally "identical" with an original object.
with the specificative sentences that cannot be significantly expressed in terms of identity over time for objects like ourselves.

For instance, imagine a case like Parfit's Combined Spectrum, in which both physical and psychological connectedness change gradually. At the beginning of the spectrum, ("just a few steps ahead!!") the person evolving is both psychologically and physically similar to the original person, but at the very far end of the spectrum, the person reached is neither physically nor psychologically similar to the starting person. If I were to enter such a Combined Spectrum, I could ask myself a question such as "If I take this further step, would then the evolving person be me, or would the change occurring imply that I would die in that case?" At the beginning of the spectrum, the answer to such a question would certainly be that I would be that very person, since the evolving person would to a large extent be psychologically and physically continuous with me. The same argument would show that at the very end of the spectrum the evolving person would not be me. If I then ask that question somewhere in the middle of the spectrum, would it not be the case, claims the Indeterministic Criterianist, that I do not know how to answer such a question about my survival. Not because I do not know what happens in the middle of the spectrum, since I would really know to what extent I would be psychologically and/or physically continuous with the resulting person, but because the level of description involving 'identity' cannot capture what actually takes place. It seems to be the case, then, that I do know everything there is to know about the case of the combined spectrum, but still not know whether identity obtains. That is, not know if I am about to die. The Indeterministic Criterianist says that it would be an empty question whether I survive, or whether there is a completely new person created. There can be empty questions about our identity over time, just as there can be empty questions about the identity over time of artefacts, since personal identity over time consists in a relation that is not an all or nothing relation, but a matter of degree.

4.9 The reductionist thesis

Some philosophers are Non-reductionist Criterianists and uphold the idea that the specified conditions cannot be specified without presupposing the fact that individuals like us exist. Other philosophers, such as Derek Parfit, are Reduc-

---

90 *Ibid.* pp. 213-217,
tionist Criterianists, since they claim that it is possible to specify what our identity over time consists in without presupposing that persons exist. Still others, are Strongly Reductionist Criterianists since they uphold the idea that 'personal identity' and 'person' do not refer to any fact whatsoever and that persons and personal identity are purely fictional entities.91

A Non-reductionist Criterianist endorses the idea that statements about our identity over time can be true if and only if some appropriate statement or statements about the psychological and/or physical continuity are true. In this sense, our identity over time invokes nothing over and above the existence of psychological and/or physical continuity just as a nation invokes nothing over and above the existence of its citizens. But, an important point with this idea is that it does not purport to eliminate reference to persons when it gives the necessary and sufficient conditions of our identity over time. It is even claimed that giving the logically necessary and sufficient conditions for personal identity over time is to give an informative account of the "diachronic criterion of personhood",92 i.e., through specifying facts about psychological and/or physical continuity we establish the relation that must obtain if two persons are one and the same person. That is, even though the Non-reductionist Criterianist purports to give weakly impersonal necessary and sufficient conditions for personal identity over time, they do not purport to eliminate all talk about persons. On the contrary, a Non-reductionist Criterianist explicitly claims that the meaning of 'person' is distinct from there being a brain and a body and/or mental experiences, and that the specification of the necessary and sufficient conditions of personal identity over time needs to be stated as facts about persons. This is because the concept of someone being in a mental, or bodily state, is prior to the concept of a mental or bodily state.93 The reason for this is that they are states of persons. Claiming that mental states and bodily states are states of persons makes it possible for a Non-reductionist Criterianist to consistently claim that the determination of truth-conditions of the specificative class statements about psychological and/or physical continuity even presupposes the identification of a person as the owner of those mental, and bodily states. It is not until we have carried out this identification of the person that we can give the necessary and sufficient conditions of personal identity over time.

92 Noonan, Personal Identity, p. 111.
93 Ibid. p. 87.
A Non-reductionist Criterianist would claim that what we mean by 'person' is different from the meaning of brain and body, but that personal identity over time implies the existence of a brain and a body. If we take the statements about the psychological and/or physical continuity to be the specificative class, we can say that the statements about the psychological and/or physical continuity are a guarantee for the truth of the corresponding statement about personal identity over time, and that the understanding of statements about personal identity over time implicitly include some knowledge of those statements relation to statements of psychological and/or physical continuity. An analogy would be the purported specification of facts about nations into facts about people. Even if a nation is nothing over and above the people at a certain place and at a particular time, it was France that declared war on Germany in 1939, not the people living in France. In the same way that we do not reduce talk about nations into talk about citizens we do not reduce talk about personal identity over time into talk about psychological and/or physical facts. It is persons who love, hate, walk and not their brains, bodies or experiences. That is, persons are distinct from their brains, bodies and experiences in the same way as a nation is distinct from its citizens.

Reductionist Criterianism, on the other hand, claims that the specification of necessary and sufficient conditions for personal identity over time can be specified in a strongly impersonal way. As we just saw, Non-reductionist Criterianists claimed that both mental and bodily states had to be states of persons, but the Reductionist Criterianist thinks that it is possible to talk about experiences, and fully describe the relation between them, "without claiming that these experiences are had by a person".

Because of the requirement of strongly impersonal description, the Reductionist Criterianist has to work out two specifications, or reductions, in giving the informative account of personal identity over time. The first requirement is to specify the facts that constitute a person at a time without presupposing the concept person. According to Reductionist Criterianism, this is done through redescribing mental states and bodily states as being events. Events, contrary to states, need not be of something. Given this reduction of states into events, it

---

94 Parfit, Reasons and Persons, p. 472.
95 Noonan, Personal Identity, p. 87.
96 Parfit, Reasons and Persons, p. 212.
97 Ibid. p. 217.
98 Ibid. p. 211.
is possible for a Reductionist Criterionist to claim that the specification of personal identity should have the form:

If \( a \) and \( b \) are experience events and/or bodily states, then the person of which the event \( a \) is a manifestation at \( t \) is one and the same person as the person of which the event \( b \) is a manifestation at \( t' \) if and only if the condition \( C \) holds between \( a \) existing at \( t \) and \( b \) at \( t' \) (\( \langle a, t \rangle C \langle b, t' \rangle \)), where \( a \) and \( b \) can be described without mentioning the existence of a person.\(^9\)

In this sense, there need not be any reference to persons in the ultimate analysis of personal identity. Persons are objects of a secondary or derivative category.

But, even though we could give this impersonal description of the identity over time of a person, this is not something that must be done according to the Reductionist Criterionist. We could choose to describe the same fact in a terminology that presupposes the existence of persons, and in which mental and bodily states are states of persons. The reason that we do not need to give an impersonal description is because persons really exist. But, the existence of persons is a fact which could be described in several different ways, just as we can describe the planet Venus as Venus or the morning star.\(^10\) But, the real existence of persons is, according to the Reductionist Criterionism, only real in a conceptual sense, so that it is possible to change our conceptual scheme if we find it necessary, in such a way, as not being committed to claiming that persons exist. It is even claimed that we should repudiate those entities which we have no reason to believe in, and since the Reductionist Criterionist does not think that there are any good reasons to believe in persons, we should reject this belief.\(^11\) As Parfit says;

In explaining the unity of this life, we need not claim that it is the life of a particular person. We could describe what, at different times, was thought and felt and observed and done, and how these various events were interrelated. Persons would be mentioned here only in the description of the content of many thoughts, desires, memories, and so on. Persons need not be claimed to be the thinkers of any of these thoughts,\(^12\)

and


\(^11\) Ibid. p. 224.

\(^12\) Ibid. p. 251.
Though persons exist, we could give a complete description of reality without claiming that persons exist.\footnote{Parfit, \textit{Reasons and Persons}, p. 212.}

Strong Reductionist Criterianism also holds that it is possible to give a complete description of personal identity over time without ever mentioning the existence of persons. While the Reductionist Criterianist claims that persons have a conceptual existence, so we still can talk truly about persons as existing over time, the Strong Reductionist Criterianist claims that persons simply do not exist. The most famous Strong Reductionist Criterianist is David Hume. According to Hume, what we call 'persons' are nothing but "fictions" created by the human imagination,\footnote{Hume, \textit{A Treatise of Human Nature}, p. 259.} and since the term ‘person’ does not refer to any real object we cannot truly talk about personal identity over time either.

The difference between Strong Reductionist Criterianism and Reductionist Criterianism is a question of attitude. While the Reductionist Criterianist only claimed that it was possible to give a specification, or reduction of personal identity over time in an impersonal way, the Strong Reductionist Criterianist claims this is something that we ought to do if we want to speak truly about the world. Persons, truly considered, do not really exist. As Hume said,

> For my part, when I enter most intimately into what I call \textit{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 \textit{myself} at any time without a perception, and never can observe anything but the perception,\footnote{\textit{Ibid.} p. 252.}

and that the

> identity, which we ascribe to the mind of man, is only a fictitious one.\footnote{\textit{Ibid.} p. 259.}

Since persons do not exist, neither does personal identity over time. The correct view of what we falsely call personal identity over time is a bundle of a bundle of events, since persons consist of nothing more than the psychological and/or physical continuity expressed by the necessary and sufficient conditions. The things that we improperly call persons are nothing but a series of interrelated physical and/or mental events.
4.10 Non-Criterianism

Non-Criterianism strongly opposes some, but not necessarily all of the central claims of Criterianism. What all Non-Criterianists have in common is that they cannot accept the specifiability thesis and the derivativeness thesis.

According to Non-Criterianism, our identity as persons is something primitive in relation to the empirical continuities that might underlie it. The primitiveness of personal identity is shown by the fact that we cannot specify what it is, or what it means, to be one and the same person over time in a non-informative and non-circular way. That is, we can only specify what it is to be the same person over time by involving or presupposing personal identity over time. In this sense, according to Non-Criterianism, personal identity is an unanalysable fact. Personal identity is a primitive fact which cannot be eliminated, or reduced into some more basic facts. Because personal identity resists analysis it is said to be a primitive "further fact".

But, Non-Criterianism does not need to oppose the essentiality thesis, nor the complexity thesis. It does not need to oppose the essentiality thesis on the ground that a Non-Criterianist can accept that the empirical continuity specified by a Criterianist would, if it captured necessary and sufficient conditions for personal identity, really be a description of what essentially is involved in a case of personal identity. But, the Non-Criterianist continues, the fact that no such necessary and sufficient conditions can be informatively specified shows that personal identity is essentially something other than the empirical continuities specified in the specificative class.

Non-Criterianism can very well incorporate the fact that personal identity over time is something complex, since a Non-Criterianist can uphold the view that it is possible to specify necessary, but not necessary and sufficient conditions for personal identity over time. This implies that a Non-Criterianist denies the possibility of establishing determinate conditions for the beginning to exist or ceasing to be for persons, but accepts that it is possible to find conditions constraining personal identity in the sense that it is possible to know that certain conditions are necessary for personal identity.107 These necessary conditions could, then, determine that personal identity over time is not something simple, since it is partially dependent upon the continuation of some more basic facts, such as, for instance, some form of psychological and/or physical continuity.

In short, the central claims of Non-Criterianism are:

---

(i) The unspecifiability thesis, i.e., it is not possible to specify personal identity in terms of necessary and sufficient conditions

(ii) The primitiveness thesis, i.e., personal identity over time is primitive in relation to the empirical continuities that might underlie it.

But, just as there exists several different versions of Criterianism, so there exists different versions of Non-Criterianist theories. The reason that Non-Criterianist theories differ from each other is that they give different accounts of why personal identity over time is primitive, and what this primitiveness of personal identity consists in.

In distinguishing between different Non-Criterianist theories, I will use two different criteria:

(i) the account given of why personal identity over time is unanalysable and primitive,

(ii) the account given of what this primitiveness of personal identity consists in.

Given these criteria we can distinguish between Non-Criterianists who claim that personal identity is primitive on the ground that our subjective first-personal perspective gives us the fundamental and essential properties of persons. That is, the term 'I' and the referent of 'I' is understood as the primitive unit by which our knowledge of the primitiveness of personal identity and persons can be established.

Contrasted with this pure subjective perspective, with its focus upon the first-person, is a view which not so much focuses upon our first-person perspective, though it also incorporates insights given by that perspective, but a view which claims that personal identity is primitive on the ground that we have to understand ourselves as persons, and that by reflecting upon the sortal concept person we can establish certain necessary truths about ourselves. That is, the sortal concept person is understood as the basic primitive unit by which our knowledge of the primitiveness of ourselves and our identity over time can be grounded. Non-Criterianism based upon the subjective, first-person perspective has its historical root in the writings of Descartes, and has been advocated in more recent time by Roderick M. Chisholm, Richard Swinburne and Geoffrey Madell. Non-Criterianism based upon reflections on the sortal person is very much centred around

views once expressed by P. F. Strawson, (though Strawson himself was not a Non-Criterianist), and in contemporary philosophy by David Wiggins. A hybrid version, though mainly focusing upon the subjective perspective, has been advocated in recent times by E. J. Lowe.

The second criterion, viz., the account a Non-Criterianist gives of what the primitiveness of personal identity consists in, is tantamount to an elucidation of what kind of entity a person is. I will distinguish between Non-Criterianist views which identify persons with some form of psychological entity and Non-Criterianist views which identify persons with animals.

---


110 Lowe, *Subjects of experience*.
5. PSYCHOLOGICAL CRITERIANISM

5.1 Introduction

In this chapter I will begin examining particular theories of personal identity, and I will, ultimately conclude with a refutation of the Criterianist position. In this chapter I will be focusing upon Psychological Criterianism. In the next chapter I will discuss Physical Criterianism, and Animalism will be investigated in chapter 7.

Psychological Criterianism has been advocated, in one form or another, by philosophers such as John Locke, H. P. Grice, John Perry, Sydney Shoemaker and Derek Parfit. What all of those theories have in common is the idea that, even if it is the case that some kind of physical states are necessary for being a person, it is the unity of consciousness which is of decisive importance for personal identity over time. In this sense, ‘person’ is a term which picks out a psychological, or mental, “thing”. In claiming this, all Psychological Criterianists entail the view that personal identity consists in the continuity of psychological features.

I will distinguish between three forms of Psychological Criterianism. On the one hand, an account which claims that personal identity consists in continuity of memory. The second account claims that personal identity consists in the continuity of several different psychological states or features. The third account...

---


is an elaboration of the second form and it maintains that personal identity consists in the continuity of quasi-psychological states.\(^8\)

I will argue against all of these accounts. This I do by using the circularity objection originally created by Joseph Butler and Thomas Reid against John Locke's memory criterion of personal identity. I will show that Psychological Criterianism is an adequate account of personal identity only if the idea of there being quasi-psychological capacities is correct.\(^9\) But, I firstly claim that quasi-psychological terms cannot be specified without presupposing personal identity on the ground that they are causal terms.\(^9\) Secondly, the informational content that is said to be in a quasi-memory implies that it is impossible to specify the having of a quasi-memory without presupposing the concept *personal identity*. 

A third argument focuses upon the question whether the having of a quasi-psychological experience, such as a quasi-memory, is intelligible for a person. If, as I think, this is impossible, we have a reason for thinking that talk of quasi-psychological faculties makes no sense and that it has no significant function in an accurate account of personal identity. I will also argue that Psychological Criterianism cannot be upheld on the ground that the only possible way in which a mental state, be it an ordinary mental state or a quasi-psychological mental state, can be individuated is through individuating it in relation to the person who is having the mental state. This implies, then, that personal identity is prior to the succession of mental states, and hence that personal identity cannot be specified in terms of psychological continuity.

These four arguments provide, I think, a strong case against any form of Psychological Criterianism and I conclude that Psychological Criterianism is inadequate as an account of personal identity.

---

\(^8\) Sydney Shoemaker, 'Personal Identity: a Materialist's Account', from *Personal Identity*, (Oxford: Basil Blackwell Publisher Limited, 1984). See especially Parfit, *Reasons and Persons*. Parfit's idea is that all psychological notions, and not only memories, could be characterised in terms of this "quasi-" notion, pp. 260-261.


\(^1\) In the following, I will mainly deal with the notion of quasi-memory. This is because most of the discussion has been focused on that idea, and it will shorten my account of the criticism. So the choice of criticism levelled against the notion of quasi-memories should be understood as a choice governed by pragmatic reasons only, and it should be understood as equally directed to other kinds of quasi-psychological features and faculties.
5.2 Locke's theory of personal identity

According to Locke, a person must be viewed as

a thinking intelligent being that has reason and reflection, and can consider itself
as itself, the same thinking thing, in different times and places....

By pressing on this mental capacity of the person, and by claiming that both
the metaphysical principle of individuation (i.e. existence itself) and that of
being a thing of a determinate kind is needed for an adequate principle of iden-
tity for a specific object of a kind, Locke was able to sharply distinguish the
idea of 'person' from the idea of 'man' and soul. Since those ideas are of differ-
ent kinds of entities, and being an object of a certain kind is a part of the prin-
iple of identity, Locke thought that there could be identity of soul and identity of
man, without there being anything close to personal identity. The principle of
identity for persons should account for our idea of 'person'. Since the idea of
'person' is that of a thinking object which can consider itself as the same think-
ing object in different times and places, and since a thinking entity can con-
ceive itself as being at different times and places only if it is consciously aware
of mental activities had in the past and at the present time, Locke thought the
faculty of reflexive consciousness is what enabled a person to conceive himself
as one and the same person existing at different times. Concentrating on this
awareness of mental activities given in consciousness, Locke required that a cri-
terion for personal identity should be conditioned by only one thing: it should
imply and take account of the importance we attach to the unity of conscious-
ness.

The most obvious explanation of this unity of consciousness is an account
which focuses upon the vivid memories of earlier episodes that a person has,
since the faculty of memory provides persons with an exceptionally good way of
gaining knowledge of the past. Since the knowledge given through the faculty of
memory is from the inside it gives persons

---

94 Locke, An Essay concerning human understanding, Book II, Chap. XXVII, p. 342. It is
imaginable that there exist a person who claims that he has the same soul as Socrates, but
that the person in question does not remember anything of the deeds done by Socrates. In
that case, according to Locke, we could not say that the latter person is identical with
Socrates, although we could claim that the latter persons soul is identical with Socrates soul.
95 Ibid. p. 335
96 Ibid. p. 335
Let us call memories which are presented to the person from the inside 'personal memories'. Personal memories enable a person to remember experiences and events as being parts of his own history, that is, as being episodes in the same consciousness.

An important point with Locke's account of personal identity is that the personal memories which constitute personal identity are apparent memories. When I have a memory-experience of myself fishing in a stream in the northern parts of Sweden, the having of that experience is independent of whether it is true or not that I actually once went fishing in the northern parts of Sweden. My memory-experience is, in this sense, an apparent memory. In this sense, an apparent memory need not correspond to any actual experienced event or action of the past. It only reports what a person seems to be remembering from his own perspective. Since the apparent memory does not need to be true, it is only a memory in a weak sense. Contrasted with apparent memories are veridical memories. I have a veridical memory of an event E if and only if the remembering of E corresponds to the earlier experience of E. In this sense, a veridical memory, contrary to apparent memories, must be true, and should therefore be regarded as memory in a strong sense. In claiming that personal identity over time is constituted, or consists in, the ability to have memory-experiences of the earlier existence, Locke was thinking of a faculty of memory which provides a subject with apparent personal memories with their associated immunity to error through misidentification relative the subject concerned in the memory.

According to Locke, when a person has a present conscious memory-experience of an event E, or an action A, he identifies himself with a particular past conscious experience of that event E, or action A. The apparent memory of that past conscious experience establishes that it is the same consciousness that is doing the present conscious remembering, and which had the conscious experience of E. This means that the unity of consciousness which is necessary for personal identity is provided by the faculty of memory, and that Locke's account of personal identity is that personal identity consists in memory. Given this, Locke's theory of personal identity can be stated as follows:

A later person P* is the same person as an earlier person P if and only if the later person P* is able to have memory-experiences of the attitudes, actions and events witnessed by the former person P.

---

The plausibility of this view is grounded in the fact that reports about personal memories are immune to error through misidentification relative to the first person pronoun. Sydney Shoemaker characterises the possibility of error through misidentification as

a statement ‘a is F’ is subject to error through misidentification relative to the term ‘a’ means that the following is possible: the speaker knows some particular thing to be F, but makes the mistake of asserting ‘a is F’ because, and only because, he mistakenly thinks that the thing he knows to be F is what ‘a’ refers to.98

That is, if we can express knowledge that someone is F, without knowing whom it is that is F, then our statement is subject to error through misidentification. To gain knowledge about the truth of such a statement we need to perform two different acts; first we need to establish that something x is an F, and then we have to identify that that x which is F is identical with an object a. The statements whose knowledge is in this sense dependent upon two acts of recognition, Gareth Evans has called, “identification-dependent”99 But, certain kinds of statements are not thus identification-dependent. If there is no difference between having the information that Φ, and the having of information that it is a that is Φ, then a person need not identify x with a, since the person's disposition to have the judgement “a is Φ” is completely determined by what object it is in his field of experience that gives rise to the judgement in the first place. This means that if it is necessary for knowing the truth of a statement a is F, that a is the source of the information that something is F, then the statement a is F is immune to error through misidentification. As Evans formulates it, in a judgement which is immune to error through misidentification it

does not appear to be a gap between the subject's having information (or appearing to have information), in the appropriate way, that the property of being F is instantiated, and his information (or appearing to have information) that he is F; for him to have, or appear to have, the information that the property is instantiated just is for it to appear to him that he is F.100

A criterion for a statement to be immune to error through misidentification relative to the first-person pronoun is that it is meaningless to ask a question such as “Someone did F, but was it I who did F” because knowing the truth of F is only possible through knowing that it was oneself who did F. For instance,

typical statements that are immune to error through misidentification relative to
the first person pronoun are statements attributing P-predicates of the subject
where the knowing that the P-predicate is attributable to someone is the same as
knowing that it is attributable to oneself. For instance, consider the P-predicate
‘pain’. In making the assertion ‘I am in pain’, the information justifying the
attribution does not differ from the information justifying that it is I who am the
subject who is having a pain.

Another example that brings the central idea of immunity to error through
misidentification relative to the first person pronoun into focus is a judgement
such as ‘I see a tree’. In a very simple sense this judgement is not immune, since
we can easily imagine that I am making a false statement in claiming ‘I see a
tree’. I might be drunk, or be hallucinating, and not see a tree at all, even though
I believe I do. But, even though it is true that I might be wrong in thinking that I
see a tree, I simply cannot be wrong concerning that it is I who am perceiving
something which looks like a tree, given the information leading up to my mak­
ing the judgement in the first place. That is, in statements which are immune to
error through misidentification relative to the first person pronoun one cannot
mistakenly think the P-predicate one is ascribing to the subject is true of some­
one else than oneself.

Statements or judgements attributing personal memories to a subject seem to
satisfy this criterion for being judgements which are immune to error through
misidentification relative to the first person pronoun. For instance, to have the
information that a particular memory is instantiated is nothing but it appearing to
the subject that he is having the memory. But, the immunity to error through
misidentification of first personal memories comes in two different senses.101
Firstly, the knowledge given in inside remembering would be knowledge-pre­
serving in a very special sense. If I seem to remember seeing Paul at the railway­
station, it is possible that I misidentify my memory as being of Paul, since it
could have been Paul's twin-brother John that I saw at the railway-station in the
first place. But it cannot be the case that I have a memory of a thing X, and at
the same time be uncertain that it was I who witnessed the thing X. Secondly,
the belief structure in a first-personal memory content is belief-preserving when
it comes to the ‘T’. If I seem to remember seeing Paul at the railway-station, it is
possible that I should, either at the time of observation or at the time of the
remembering think myself to have misidentified a person as being Paul, and that
my memory of the episode would not be accompanied by the belief that it is a

memory about Paul. But it cannot be the case that I have a memory of a thing X, and at the same time not have the belief that it was I who witnessed the thing X.

5.3 The circularity-objection

According to Locke a person’s identity over time is constituted by, or consists in, the apparent memories that are present in a consciousness. As long as one person remembers certain experiences and actions, that person is the same as the one who underwent the experience or did the action.

The natural objection to this analysis is that it is viciously circular, and that we cannot define personal identity over time in terms of memory, since the concept of memory implicitly invokes the conception of personal identity. As Joseph Butler once formulated it;

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. 102

The central point of this objection is that it focuses on the fact that Locke, in his specification, has put the matter in completely the wrong order. The objection acknowledges, as Locke also did, that there is a very close connection between personal identity and memory. It is by the faculty of memory that we have access to our earlier experiences. But, the objection goes, we cannot specify personal identity in terms of memory as Locke tried to do, since the only way of understanding memory is by presupposing the identity between the one who is remembering and the one who had the experience. The reason that we cannot define personal identity in terms of apparent memories is that apparent memories are accompanied by a “previous awareness condition” in which the content of an apparent memory already entails the identity over time of the person who had an experience in the past with the person who is presently remembering the occurrence of the experience in the past. 103 This can be shown in an analysis of the conditions that must be satisfied in a definition of what it is for a person S to remember a certain past event E. S remembers E if and only if it is the case:

(1) that S now has a state (that could be dispositional) which could be called an apparent memory, and

(2) that the content of that apparent memory “matches” in an appropriate way the nature of the past event E, and

102 Joseph Butler, ‘Of Personal Identity’, p. 100
103 Shoemaker, ‘Persons and their Pasts, p. 269.'
(3) that S was appropriately related to E at the time of its occurrence.\textsuperscript{104}

The previous awareness condition shows itself in condition (3). As condition (3) implies, if S is to remember E, then S must have been around when E occurred. The fact is that true memories, or veridical memories, require that it is one and the same person who is doing the remembering and who had the experience, since memories require a causal relationship between E and the remembering of E. If it had not been the case that the person who witnessed the event and the person who remembered the event was one and the same person, we could not claim that S was remembering the event E, but rather that S had an illusion as of remembering E. In this sense it is the case that remaining the same person is a pre-condition for having memories. The reason why I can remember my first day at the university is that I am identical with the person who went to the university on that particular day. It is not, as Locke would have to say, that I am identical to the person who went to the university on that particular day because I, today, can remember that particular occurrence. In this sense, personal identity over time is semantically-\textit{cum}-ontologically prior in relation to memories.\textsuperscript{105}

The circularity objection is not the only line of criticism levelled against Locke's original version of Psychological Criterionism. The most easily answered objection for a Psychological Criterionist is that of the failure of transitivity by Thomas Reid.\textsuperscript{106} This objection is essentially that identity must be transitive, in the sense that if P is identical to P*, and P* is identical to P', then P is identical to P'. But this seems not to be the case with memory. If P* remembers being P, and P' remembers being P*, then it does not follow that P' must remember being P. P' could have forgotten all the things that happened to P. According to Locke's analysis, then P' is not identical to P, but P* is identical to P, and P' is identical to P*. The answer to this objection is to take the succession

\textsuperscript{104} Shoemaker, 'Personal Identity: a Materialist's Account', p. 82

\textsuperscript{105} But, this does not mean that memories are uninteresting from the point of personal identity over time. Firstly, that I can remember a certain thing A that occurred in the past, does in fact, give me evidence that I was around at the time of the occurrence of A. That is, the special kind of access of our lives given by the faculty of memory is interesting from an epistemological point of view. But of course this evidence is fallible. My memory of A could have been induced in me by a hypnotiser, so that I falsely believed that I was around when the occurrence of A took place. Not knowing that A is an illusion A would still be one of the determinations of my psychological conception of who I am. In this sense, memories can never be anything but fallible evidence of what or who a person is. Secondly, memories are also important for a person's self-understanding. Memories enable a person to have access to earlier periods of his past through recollecting past experiences which to a large extent were part of a process of self-creation in which the person's self-knowledge and self-conception were formed. Memories can, in this sense, tell an individual person who he is, but not what he is.

of experiences and the causality of memory into the account. David Hume\textsuperscript{107} was the first to emphasise these ideas. Relying on his ideas about the person being a bundle of bundle of perceptions, and the fact that these momentary bundles of perceptions stand in the relation of cause and effect to each other, it is possible to define two different types of memory-relations, \textit{direct memory-connection}, and \textit{memory-continuity}.\textsuperscript{108} For there to be a direct memory-relation between $P^*$ and $P$ it is required that $P^*$ should remember almost everything that happened to $P$. For there to be memory-continuity between $P'$ and $P$, on the other hand, it is only required that there should be an overlapping chain of direct memory-relations between $P'$ and $P$. What is now required for personal identity is memory-continuity, not direct memory-connections.

Another argument levelled against Locke could be called the \textit{epistemic-objection} and it has focused on the fact that Locke laid too much stress on the epistemic importance of the notion of memory. To Locke, memory is both necessary and sufficient for personal identity, since he thinks the faculty of memory is the only thing which explains the unity of consciousness. In this it seems that Locke was getting towards a correct understanding of memories as being our primary source of gaining knowledge about our past. But, it cannot be correct that memories are the only constitutive feature of the person, since it is conceivable that a person can survive in an amnesiac state.

This objection can be answered by qualifying Locke's original account through accepting that memories are not the only important feature in personal identity. Other psychological factors, like intentions, emotions, beliefs, character-traits, need to be accounted for in a correct specification of what is necessary and sufficient for remaining the same person over time. The more recent versions of Psychological Criterianism have acknowledged that it is the whole range of psychological features which are of importance for personal identity over time and also that it is the continuity of psychological features, which constitute personal identity over time, That is, it is psychological continuity, i.e. continuity of intentions, emotions, memories, beliefs etc. that constitutes the persistence of a person over time. Personal identity over time consists in the preservation of all kinds of different psychological traits of persons. This elaborated version of Psychological Criterianism can be formulated as:

(i) A person $P_2$ at $t_2$ is identical with a person $P$ at $t$ if and only if $P_2$ is psychologically continuous with $P$,

\textsuperscript{107} Hume, \textit{A Treatise of Human Nature}, pp. 251-263.

\textsuperscript{108} Parfit, \textit{Reasons and Persons}, p. 205.
(ii) that there is no person P* existing at $t_2$ who is psychologically continuous with P to the same extent as $P_2$, and

(iii) the psychological continuity has the right kind of cause.

I will not directly consider, nor directly criticise, the latter two clauses of the definition given by Psychological Criterianism. The reason for this is that I think that the circularity objection is as fatal to more recent versions of Psychological Criterianism as it was to the original version given by Locke, even though Psychological Criterianism thinks it to be the continuity of a whole cluster of psychological features which constitute personal identity. Furthermore, we will eventually discuss matters which are related to the two latter clauses of the definition.

That the elaborated version of Psychological Criterianism is subject to the circularity objection is due to the fact that every psychological notion necessarily presupposes personal identity. Let us by a ‘personal intention’ understand an apparent intention of a subject S where the apparent intention contains the goal G and G is realisable by S himself. If we, for instance, analyse the conditions that must be satisfied in a definition of what it is for a person S to have a personal intention, we will end up with the following conditions:

1. It seems to S, that S now has an intention I, with the goal G,
2. the satisfaction of the goal G is appropriately caused by the earlier intention I, and
3. that S is appropriately related to the goal G, when G is brought about.

This latter clause of the elaborated version of Psychological Criterianism is intended to rule out the possibility of two different candidates being equally good candidates for identity with an original person. If we have a “dead race” between two candidates, i.e. a case where two candidates are equally strongly psychologically continuous with the original person, neither one of the candidates can be the original person.

The last clause of Psychological Criterianism varies between different advocates of the theory depending upon how strong the requirements of physical continuity a particular advocate of Psychological Criterianism builds into the criterion. Parfit, in *Reason and Person*, (pp. 207-209), distinguishes between three different “versions” concerning what could be the right cause in a criterion for personal identity over time. According to the “narrow version”, the cause of psychological continuity constituting personal identity over time must be the normal cause, i.e. the brain and the psychological functions governing the brain. The second version is the “wide version” which accepts as a cause for psychological continuity any reliable cause, while the third, the “widest version” accepts psychological continuity founded upon any cause whatsoever.

By putting in other psychological concepts, like wishes that and dreams that, we would get the same analysis.
It is still condition (3) that makes the analysis circular. For a person to be intending an action implicitly invokes the idea that the person will benefit from that particular action, i.e. that intending relies on an “anticipating awareness condition” where the intending subject imagines a future time in which he, that very same subject, will either do, or benefit from an action. This means that we can formulate the circularity objection as claiming that it is impossible to specify personal identity in terms of psychological continuity, since psychological continuity presupposes personal identity. That is, personal identity is what grounds psychological continuity, not psychological continuity that grounds personal identity.

5.4 Quasi-memories and personal identity over time

Recent advocates of Psychological Criterionism have accepted the circularity objection levelled against all kinds of psychological features. We simply cannot specify personal identity in terms of ordinary psychological notions. But they claim, it is still possible to retain the spirit of Locke, since it is possible to construe cases in which our ordinary psychological terms no longer are seen as referring to ultimate faculties by which we can recollect, intend, and experience past or future episodes in our lives. If we still confine ourselves to memories, Psychological Criterionists no longer view memory to be the ultimate faculty by which we recollect earlier episodes in our lives. The idea is that it might be the case that we are able to gain informative knowledge of the past without satisfying the “the previous awareness condition” through a faculty of quasi-memory. Sydney Shoemaker, who was the first to propose this idea, claims that Locke’s memory-theory is too narrow in the sense that he considered memories to be the only possible domain from which we could acquire knowledge of the

---

112 The immunity to error through misidentification that is exhibited in our first-personal memory reports, do not have a corresponding immunity feature in the case of intentions. It could be the case that I intend E, but because of lack of knowledge, not know that I am not the beneficiary of E.

113 Shoemaker, ‘Persons and their Pasts’. Shoemaker distinguishes between two different possible cases of a faculty of quasi-memory: a restrictive faculty of quasi-memory, and a general faculty of quasi-memory. The difference between the general and the restrictive understanding of quasi-memories is that the general understanding does not lay any restrictions upon the type of causal chain that is supposed to link the experience of E to the apparent remembering of E, while the restrictive understanding requires that the causal chain should be of the appropriate sort, i.e. it should resemble the ordinary causal chain as much as possible. In what follows, I will not distinguish between these types of quasi-memories, but treat them on a par.
past. What Locke did not realise was that it is only a contingent fact that human beings have ordinary memories, and not quasi-memories. It might be the case that conscious beings have quasi-memories and that our ordinary memories are nothing but a sub-class of the possible class of quasi-memories. Memories are the class of quasi-memories that the person has of his own earlier past, and this is why memories are both immune to error through misidentification, and rely on the previous awareness condition. But this is not the case with quasi-memories. Due to the lack of a conceptual tie between quasi-memories, the previous awareness condition and immunity to error through identification, we can specify personal identity in terms of the continuity of quasi-memories. It is in this sense that quasi-memories are thought to be impersonal, either weakly or strongly.¹¹⁴ That is, it is possible to characterise the whole past of a person without presupposing personal identity over time.

One important aspect of quasi-memories is that they are apparent memories which present themselves in a subject-indeterminate form. That is, contrary to an apparent memory of an ordinary memory which is presented with a content containing a first-person mode, a quasi-memory should contain a general mode of presentation in which the apparent memory is accompanied by a someone or other content. This means that an advocate of Psychological Criterionism upholds the view that apparent memories are only contingently associated with a belief satisfying the previous awareness condition and that apparent memories consist of two separable components. The first being a general component that someone or other experienced E, and the second component being that the individual who has the apparent memory is either the subject of that experience E, or someone else. The idea is that a quasi-memory provides knowledge of lives from the inside”, without necessarily being recollections from the inside of one’s own past history.¹¹⁵ They would give a subject knowledge of how it was to be another person on a particular occasion. This means that the special access of remembering from the inside will be lost. From the fact that you seem to remember a certain past event E, you cannot conclude that you really witnessed that event. To determine whether you remember E, or only quasi-remember E, you must follow a causal chain backwards in time, and check whether you were one of the persons around at the time of E:s occurrence. Given that it is possible to distinguish between the general component and the identifying component of

¹¹⁴ See chapter 4, pp.
¹¹⁵ Parfit, Reasons and Persons, p. 221.
apparent memories, Shoemaker now thinks that we could characterise a quasi-memory through the following analysis:

1. that S now has a state (that could be dispositional) which could be called an apparent memory, and
2. that the content of that apparent memory “matches” in an appropriate way the nature of the past event E, and
3. that someone was appropriately related to E at the time of its occurrence.\textsuperscript{116}

Since the content of a quasi-memory does not depend upon a content where oneself is contained as the original experiencer, quasi-memories do not rely upon the presupposition that it was oneself who experienced the original event or did the original action that one is now having an apparent memory of. Hence, the previous awareness condition is not really a presupposition for having apparent memories.

As an example of how a quasi-memory would be given to a person, consider the following scenario imagined by Derek Parfit:

\textit{Venetian Memories}. Jane has agreed to have copied in her brain some of Paul's memory-traces. After she recovers consciousness in the post-surgery room, she has a new set of vivid apparent memories. She seems to remember walking on the marble paving of a square, hearing the flapping of flying pigeons and the cries of gulls, and seeing light sparkling on green water. One apparent memory is very clear. She seems to remember looking across the water to an island, where a white Palladian church stood out brilliantly against a dark thundercloud.\textsuperscript{117}

Jane's apparent memory would in this case not be an ordinary memory, nor an illusion according to Parfit.\textsuperscript{118} The fact that Parfit does not take this to be an illusion needs a little more examination. In connection with Locke's memory-criteria of personal identity I distinguished between two different notions of memory, i.e. apparent, and veridical memory.

A memory-illusion can be understood as an apparent memory that is not a veridical memory. In an illusion we have an apparent memory, but this apparent memory does not correspond to, nor is it appropriately causally related to an experience of the past. If I, for instance, as a result of being hypnotised, have an apparent memory of crossing the Rubicon, and this apparent memory, as a mere coincidence, happens to be qualitatively indistinguishable from Julius Caesar's experience when he crossed the Rubicon, this would be an illusion. That is, it would be an illusion of mine, not a memory of mine. What this comes down to

\textsuperscript{116} Shoemaker, 'A materialists account', pp. 81-82. My emphasis.

\textsuperscript{117} Parfit, Reasons and Persons, p. 220.

\textsuperscript{118} Ibid. pp. 220-223.
is that the absence of a causal connection between the experience and the apparent memory deprives the apparent memory of its' genuineness, so that the apparent memory would be an illusion if it were not causally related to the experience, even though someone actually did experience a situation that qualitatively corresponds to the apparent memory. That this is not the case in Parfit's example of Jane's apparent memories of Paul's experiences, seems to be pretty clear from the description of the case. As a matter of fact, there is a direct relation between Paul's experience and Jane's apparent memory of that particular experience, and since the apparent memory of Jane would be causally connected to the experience had by Paul it is not the case that Jane's apparent memory of Paul's experience is a memory-illusion. The only way to understand this kind of case seems to be that Jane is having a veridical quasi-memory of Paul's experience.

Parfit also thinks it is possible for Jane to know that she is having a quasi-memory. She could know this through her knowledge that she has never been to Venice, but that Paul frequently travels to that town, and she knows that she has gone through an operation in which some of Paul's experiences were implanted in her brain. So Parfit thinks that after Jane has gone through the operation, she could no longer conclude that if she has an apparent memory of an event E, then she must have been witnessing E. All she can conclude is that someone or other witnessed E, and then identify that someone or other with herself or someone else. If she can identify herself with that someone or other, her apparent memory would be a veridical memory, and if she cannot identify herself with that someone or other, her apparent memory would, at best, be a quasi-memory.

That it is possible that we can have apparent memories containing the general component and the identifying component has been argued on the ground that the causal chain leading from an experience of an event to the recollection of that event can divide. If it is the case that the causal chain branches, my use of 'I' at the time of the recollection no longer can be understood as referring to the same person as my use of 'I' before the branching, since we in such a case would have two distinct persons recollecting the same event on the same causal ground. This means that our first personal reports of our apparent memories no longer need to be immune to error through misidentification. The use of 'I' in a report like "I smashed the window" could be mistaken, since the causal chain linking the event to the remembrance of the event could have branched at an

---

intermediate point between the smashing of the window and the recollection of smashing the window.

Another line of argumentation for the possibility of separating the having of an apparent memory of \(E\) from the belief that you were around at the time of the occurrence of \(E\) has been put forward by Harold W. Noonan. Noonan tries to establish that the informational content of apparent memories is determined by what belief-contents they have as a result, and not their intrinsic features. Noonan characterises a possible world where the general faculty of quasi-memory would be a natural phenomenon in that children spontaneously report information from their parents’ lives. Those reports are not reported in a first-person form by the children due to the linguistic training of the children. The first-personal form is only used when there is evidence of the fact that the quasi-memory is a genuine memory. In this world, Noonan thinks, the content of apparent memories of \(E\)-ing need not be equivalent to having the belief that it is oneself who is \(E\)-ing, since the content of the apparent memories that are quasi-memories are not given in the first-person. The only thing necessarily true of apparent memories, according to Noonan, is that they are from a particular point of view. In this sense, the content of an apparent memory need not involve the identity of the subject in a similar way as the identity of the subject need be no part of what is imagined when one imagines \(F\)-ing.

Psychological Criterianism assumes that whatever holds for quasi-memories holds for every other type of psychological faculty. As Parfit claims:

> It may be a logical truth that we can intend to perform only our own actions. But we can use a new concept of quasi-intention. One person could quasi-intend to perform another persons actions. When this relation holds, it does not presuppose personal identity.

With the help of those quasi-psychological notions, we could according to Parfit, give an account of personal identity over time that is not susceptible to the circularity-objection, since we can describe psychological continuity in an impersonal way.

(i) A person \(P_2\) at \(t_2\) is identical with a person \(P_1\) at \(t_1\) if and only if there is unity of consciousness between \(P_2\) and \(P_1\)

\[121\] Ibid. p. 184.
\[122\] Ibid. p. 183.
\[123\] Ibid. p. 183.
\[124\] Ibid. p. 185.
(ii) the unity of consciousness consists in the continuity of overlapping chains of quasi-memories, quasi-intentions, quasi-beliefs, quasi-wishes, and quasi-character,

(iii) there is no person $P^*$ at $t_2$ who is psychologically continuous with $P_1$ to the same extent as $P_2$, and

(iv) the psychological continuity has the right kind of cause.

5.5 Criticism of Psychological Criterianism

What I would like to claim is that recent psychological approaches to personal identity, contrary to appearance, also fail because of the circularity objection, since they really rest upon a prior understanding of personal identity.

Firstly, the quasi-psychological account of personal identity is circular because the assumed causal connection between the quasi-memory of an event $E$ and the occurrence of $E$ already presupposes personal identity. According to the idea of quasi-memory, the important relation between the having of an apparent memory $M$, and the event $E$ that corresponds to that memory, is that $E$ should be causally responsible for the occurrence of $E$. This causality relation could be understood in two different ways; either we could restrict the causality relation to be of the ordinary kind of causality, or we could claim that there are no such restrictions at all. Sydney Shoemaker does the first, while Derek Parfit and Harold W. Noonan do the second. What I would like to claim is that neither the restrictive, nor the general understanding of quasi-memories can be given without a pre-understanding of personal identity over time.

The circularity becomes apparent if we ask how one should establish that an apparent memory is a memory or a quasi-memory. According to the restrictive notion of quasi-memory, an apparent memory is a quasi-memory if

1. there is an apparent memory $M$, and
2. $M$ "matches" in an appropriate way the nature of the past event $E$, and
3. that $M$ is linked by \textit{the right type of causal chain} to $E$.

The right kind of causal chain should be the ordinary kind of causal connection that exists between a present memory-state $M$ and a past event $E$, but where the causal chain has branched. Since the two resulting products of such a

---

branching can have apparent memories of one and the same ancestor without being identical with each other they would quasi-remember the earlier experience E. The problem, as I see it, is whether it is possible to understand what it would be like to have a case of a branching causal chain without already presupposing personal identity. Given that we do have a causal chain between an apparent memory and a past event E, the only way to determine that this is a case of quasi-remembering, and not a case of remembering, is that we can establish that the apparent rememberer is an offshoot of the one who experienced E. Knowing that someone is an offshoot, and not one and the same person, requires an understanding of personal identity, since that knowledge relies upon knowing that there is another person who is also causally linked to the original person.

The same is also true of the general notion of ‘quasi-memory’. The general notion of quasi-memory is such that we have a case of quasi-memory if

1. there is an apparent memory M, and
2. M “matches” the nature of the past event E, and
3. that M is linked by any causal chain to E.

By the claim that any causal chain suffices for having quasi-memories is meant that any reliable causal source linking an apparent memory M to an experience E is sufficient for having a quasi-memory. But, as I see it, the only way in which it is possible to know that an apparent memory is a quasi-memory is by knowing that the person who had an experience E is distinct from the person who is having an apparent memory of that experience.

Furthermore, since all quasi-memory reports must be presented to the quasi-remembering person in the first-person mode of presentation it cannot be the case that apparent memories really can be separated into two distinct belief-components. As Gareth Evans claims, the mere fact of the possibility of constructing the linguistic notion of ‘quasi-memory’ does not show that it is possible that we really could have had such a faculty in the same way as the possibility of constructing the notion of ‘quasi-perception’ does not entail the intelligibility of a faculty of quasi-perception.127

127 Evans, The Varieties of Reference, p. 248. A subject is said to quasi-perceive E if

(i) the subject seems to see an E, and
(ii) E is the input for a causal process that has the seeing of E as a result,
(iii) the location of E either in space or in time need not correspond to where the subject is disposed to locate it.
If quasi-remembering is to be a plausible faculty, and not only a linguistic peculiarity, we must be able to have an informational state that resembles a personal memory state without having the belief-content that the subject himself did X.\textsuperscript{128} As we know, quasi-memories necessarily involve the existence of apparent memories, and these apparent memories present themselves in the first-person mode of presentation. But, since apparent memories are presented in the first-person mode of presentation this means that the apparent memory will have a particular mode of informational content that involves the belief that it was the subject himself that did the action he now apparently remembers. But, if this is the case, it seems to me that we cannot have a faculty of quasi-memory, since quasi-memories necessarily present themselves with an informational content involving the belief that it was the subject himself that did the action he now apparently remembers.\textsuperscript{129} That is, if a person P apparently remembers E, then his remembrance of E is accompanied by an apparent belief B that he apparently remembers himself witnessing E, i.e. that the apparent memory comes in a first-person mode of presentation. Secondly, since quasi-memories are defined as being apparent memories, then quasi-memories should also be presented to the person in the first-person mode of presentation. What this means is that if someone quasi-remembers E, the presentation of E must be in the form of a first-person apparent memory of E, and the way the person P will apparently remember E will in content be equivalent to 'I E-ed'. But, since 'I E-ed' has the same informational content as 'I am the same person as one who E-ed' we cannot define, or analyse personal identity in terms of quasi-memories, since our apparent memories already are accompanied by a belief that we are the same person who E-ed. The circularity would here be that we already have informational content, or beliefs, telling us that we are the same person who apparently remembers E, and the one who experienced E.

\textsuperscript{128} Evans, The Varieties of Reference., p. 248.

\textsuperscript{129} A similar argument against the possibility of a faculty of quasi-memory has been argued for by John McDowell in 'Reductionism and the First Person', in Reading Parfit, ed. Jonathan Dancy, (Oxford: Blackwell, 1987), pp. 238-244.
The third objection levelled against the notion of quasi-memories concerns the intelligibility of being in a mental state that would be a state of quasi-remembering, and where that mental state would be qualitatively the same as a state of ordinary remembering. If, for instance, we return to Parfit's Venetian Memories where Jane is said to quasi-remember Paul's experiences, an important feature of Jane's mental state is that it should provide her with knowledge about how it was to be Paul at the time of Paul's experience, and that this gaining of knowledge should resemble the kind of knowledge our memories provide us with in ordinary cases. That this oversimplifies our memory experiences is something that Marya Schechtman has pointed out. She claims that Parfit's visual sense-datum idea of memories distorts what really takes place in an actual case of remembering. Having a memory involves many different mental structures, and they typically present themselves to the subject in a personal narrative. A narrative in which reference to particular places, oneself, and other persons is of fundamental importance. If we have a case of transformation of memory as Parfit's Venetian Memory, we would either get a very confused Jane, since the experience she will apparently remember will present itself out of context, and without any self-understanding. Or, another possibility is that the whole of Paul's experience-context is transformed to Jane, but in that case we would no longer have a strictly identity-neutral quasi-memory. So, granted that we do have a case of quasi-remembering either the receiving person will not understand, and therefore not gain any knowledge about the information of the quasi-memory, or we will have a case where the receiver gains some information and knowledge of the past experience, but where we do not have a case of quasi-memories. In neither case are quasi-memories intelligible.

Let me end this criticism of Psychological Criterianism with another, and I think absolutely decisive, problem for any psychological account of personal identity. Any psychological account of personal identity must conceive it to be possible to individuate a mental state or a succession of mental states without referring to the person who possesses the mental states, since claiming that per-

---

130 A possible forth objection is that the notion of 'quasi-memory' is dependent upon the notion of 'ordinary memory', and that the characterisation of ordinary memories as being a sub-class of the wider notion of quasi-memories cannot be coherently held, since the only way to be able to understand what a quasi-memory is, is by having an understanding of memory. This argument consists in the idea that it ought to be possible to show that the analysis of ordinary memories as the sub-class of quasi-memories is circular, and then to claim that quasi-memories cannot be what grounds personal identity over time, since the faculty of quasi-memory is not a fundamental faculty.

sonal identity consists in psychological continuity is to claim that it is possible to specify an informative and non-circular criterion of personal identity. But, as E. J. Lowe has pointed out, this is not possible.\footnote{E. J. Lowe, \textit{Kinds of Being: A Study of Individuation, Identity and the Logic of Sortal Terms}, Aristotelian Society Series, Vol. 10, (Oxford: Basil Blackwell Ltd, 1989) and \textit{Subjects of experience}, (New York: Cambridge University Press, 1996).}

Firstly, it is not possible to individuate a particular mental state, say a headache, through the purely qualitative character of the mental state. For instance, it might very well be the case that we have two qualitatively indistinguishable experiences of headache, one being mine and the other being yours. But, of course, the marking difference between these two experiences is simply the fact that one is \textit{mine} and the other is \textit{yours}. But, individuating each headache in this way, either to you or me, is to individuate the mental state \textit{via} a certain person.

Secondly, neither is it possible to individuate a mental state through the causes and effects of mental states such that even though it might be true that you and I have indistinguishable mental states, say an experience of an oasis, it is still true that your experience has a cause different from mine and that your token experience will have a different effect from my token experiences.\footnote{Christopher Peacocke, \textit{Sense and Content: Experience, Thought and their Relations} (Oxford: Clarendon Press, 1983), pp.} For instance, the cause of my experience of an oasis might be a certain neural activity in my brain while your experience is caused by visually seeing an oasis while walking in the desert. Following Lowe, this account can be dismissed because (i) "the holism of the mental" makes it plain that the existence of a mental state requires reference to a whole cluster of simultaneous and earlier mental states which depend upon a prior individuation of a particular subject or person,\footnote{Lowe, \textit{Kinds of Being}, pp. 131-132 and \textit{Subjects of experience}, pp. 28-29. See also Schechtman, 'Personhood and Personal Identity' and chapter 5 of this work.} (ii) it involves a form of circularity,\footnote{Lowe, \textit{Subjects of experience}, pp. 27-28.} and (iii) that the individuation of there being a mental state requires a prior individuation of their being a subject or a person who has the mental state in question.\footnote{Lowe, \textit{Kinds of Being}, pp. 131-132 and \textit{Subjects of experience}, pp. 29-32.}

The argument from the holism of the mental is very much the same idea as the argument against the possibility of quasi-memories. A particular mental state \textit{M} will, in this sense, depend upon a whole cluster of other mental states of the person in such a way that the mental state \textit{M} would be different given that the clusters of mental states were different. A mental state \textit{M} receives its signifi-
cance in a personal narrative which make reference to particular places, oneself, and other persons. For instance, the intensity and duration of a mental state, say a pain, will depend upon the person's ability to focus upon features other than the pain and the person's overall history and beliefs of the cause of the pain.\(^{137}\) But, since a mental state depends upon the mental and physiological history of an individual person, it cannot be possible to individuate mental states without individuating them as being parts of "well-integrated mental economies" of distinct persons.\(^{138}\) A mental state \(M\), then, cannot be individuated, is not intelligible, without its relation to other mental states of a particular person. In this sense, persons, or subjects are individuated prior to their mental states.

The circularity of the psychological approach is due to the fact that the holism of the mental implies that causes and effects of mental events always involve other mental events. Since it is reasonable to suppose that some of the causes and effects of a mental event are other mental events, this implies that the mental events which are supposed to be (part of) the cause and effects of a particular mental event \(Me\) stand in need of being individuated via their causes and effects which will contain (at least some) mental events \(Me_1\). But, since the causes and effects which are to individuate a mental event are events which always involve at least some mental events we just cannot "individuate mental events in terms of the sameness of their causes and effects".\(^{139}\)

Finally, whether a certain state is a mental state depends upon whether there exists a person who is conscious of the state as a mental state. Consider for instance a case in which a split-brain patient is provided with the same music in a headphone. The auditory stimulation of the right ear will be exactly the same as the auditory stimulation of the left ear, and presumably, it is the stimulation of the right and left ear which are supposed to provide the causes of the hearing of the music. But, regardless whether we claim that there are two mental experiences of two distinct subjects or persons, or whether we claim that we have one mental experience of one subject or person, or even two mental experiences belonging to one single subject or person, we must first determine that the cause (and maybe the effects) succeeded in becoming mental states.\(^{140}\) But, doing this cannot consist in anything else than presupposing the existence of a subject, since the stimulation, in and by itself, would be insufficient if we did not have

\(^{137}\) Lowe, *Subjects of experience*, p. 28.

\(^{138}\) Ibid. p. 29.

\(^{139}\) Ibid. p. 28.

\(^{140}\) Ibid. p. 29.
one or several persons who were awake and whose capacity for hearing was functioning. In this sense, questions concerning the existence or non-existence of mental states are "only answerable in the light of information concerning the prospective subject of that experience"\(^{141}\) where that subject is "identifiable as the possessor of a suitable range of further mental states at the time of the stimulation".\(^{142}\) But, if we cannot determine whether a certain stimulation is the cause of a mental state without individuating a subject with further mental states, we simply cannot individuate mental states \textit{via} their causes and effects without a prior individuation of a person. But, since individuation of particular persons is prior to individuation of mental states, personal identity must be understood as primitive in relation to the psychological continuity. Hence, personal identity cannot consist in the continuity of psychological states.


\(^{142}\) \textit{Ibid.} p. 30. This is, once again, an appeal to the holism of the mental.
6. PHYSICAL CRITERIANISM

6.1 Introduction

In this chapter I will describe three different Physical Criterianist accounts of personal identity: the Brain criterion, the Physical Realiser criterion and the Body criterion. What these views have in common is the idea that personal identity over time should be understood from the "objective view of ourselves", i.e., firstly, that persons are partly complex physical objects composed of distinct separable parts. Secondly, that there exists no fundamental distinction between what it is that constitutes a person's identity over time and what it is that constitutes the identity over time of other kinds of concrete objects in the world. It is claimed, from the Physical Criterianist side, that both cases can be understood as the gradual development of an object, specifiable in terms of the sameness between individuated material phenomenon in time, such that an object $O_2$, at time $t_2$, is identical with an object $O_1$, at time $t_1$, if and only if $O_2$ is the same material object as $O_1$, where the object $O$ can be a ship, a stone, a lion or a person. According to Physical Criterianism, it is at least necessary that the physical matter of persons continues to exist between $t_1$ and $t_2$ if a person is to exist between $t_1$ and $t_2$. But, it is important to keep in mind that Physical Criterianists do not deny the fact that persons have a mental life. Just as the psychological accounts of personal identity understand persons as conscious and intentional beings who experience, perceive, think and have feelings, so do also different versions of the physical account. The main difference is that Physical Criterianism understands the existence of this conscious being and all its mental states as being dependent upon the existence of a certain material object, since the consciousness of a person must be caused by a certain material entity. Since persons are physical entities which have a mental life, we must, according to Physical Criterianism also take into account the fact that persons are conscious if we want to have a necessary and sufficient criterion for personal identity over

---

time. So, according to Physical Criterionism, while the preservation of a physical entity is necessary for personal identity, it is the preservation of a physical entity which is conscious that constitutes the necessary and sufficient condition for personal identity.

The **Body criterion** claims that personal identity requires the sameness of the body of the person, the **Brain criterion** maintains the view that personal identity requires the sameness of the person's brain and lastly, the **Physical Realiser criterion** claims that personal identity consists in the sameness of whatever it is that physically realises the capacity of mental states.

I will claim that the intelligibility of these accounts rests upon the idea that a later person is identical with an earlier person only if there exists no other person simultaneously competing for identity with the earlier person. Against this I will argue that it is a necessary principle of identity that whether \( a \) is identical with \( b \) should only depend upon the intrinsic relation between \( a \) and \( b \). A consequence of this is that Physical Criterionism cannot be understood as specifying a sufficient condition for personal identity, even though it could be the case that some version of Physical Criterionism specifies a necessary condition for personal identity.

### 6.2 The Brain criterion of personal identity

Given our modern scientific view of how the higher cognitive capacities are caused, a promising theory of personal identity is that our survival as the same person over time requires the continuity of the brain. Modern neurological and neuro-psychological science strongly indicates that every higher cognitive capacity, such as desire, remembrance, consciousness and intentionality, is causally dependent upon the existence of certain neural activities and patterns in the brain, since empirical findings of modern science strongly support the idea that it is physical processes of the brain which are the causal origin of our mental life. It can be argued that if persons are conscious beings, and personal identity requires the continuity of consciousness, and that it is the brain which is responsible for the consciousness of a person, then it must be the continuity of
the brain which carries with it everything required for personal identity. This line of reasoning has been taken by J. L. Mackie\textsuperscript{110} and Thomas Nagel\textsuperscript{111}

Mackie, for instance, holds the view that the terms ‘I’ and ‘person’ refer to whatever it is that “underlies and makes possible the co-consciousness of experience”\textsuperscript{112} and that unity of consciousness should be understood as the nominal essence of personal identity. The intuitive idea is here that consciousness, or even the unity of consciousness is to ‘personal identity’ what yellowish, soft metal is to ‘gold’. Our first contact with gold is through recognising certain superficial features of stuff which is gold, and some of those features constitute the stereotype of ‘gold’. But, what we ultimately mean by ‘gold’ is whatever real essence it is that causes actual instances of lumps of gold to have the superficial features normally associated with ‘gold’. In this sense, though we associate yellowish, soft metal with ‘gold’ this is not what gold is essentially. ‘Gold’ refers to all objects which have a certain inner constitution, we now think it to be the atomic number 78, and having that inner constitution is what being gold consists in. The same relation is supposed to hold between consciousness and the central nervous system, or the brain. We associate with the term ‘personal identity’ a certain unity of consciousness, but this does not mean that personal identity consists in unity of consciousness. Personal identity really consists in the continuity of the brain, since the inner constitution of the unity of consciousness of persons consists in “the structure of the central nervous system and the persistence of that structure through time”.\textsuperscript{113}

Nagel, expressing his view as an “empirical hypothesis” whose truth might be concealed to both oneself as well as to others, takes the same attitude towards personal identity.\textsuperscript{114} As Nagel says,

\begin{quote}
I am whatever persisting individual in the objective order underlies the subjective continuities of that mental life that I call mine.\textsuperscript{115}
\end{quote}

According to him, it is my brain which functions as the essential bearer and causal origin of my mental states and their continuity wherever such a continuity

\textsuperscript{112} Mackie, ‘Personal Identity’, p 220.
\textsuperscript{113} \textit{Ibid.} p 220.
\textsuperscript{114} Nagel, \textit{The View from Nowhere}, pp 40-41.
\textsuperscript{115} \textit{Ibid.} p 40.
The thought is that firstly, a person $P_2$ at $t_2$ would not be able to remember the experiences, intentions, desires and wishes of a person $P_1$ at $t_1$ unless $P_2$ and $P_1$ comprise the same physical entity. Secondly, since it is the brain that causes our mental activities it is in a determinate relation to the brain that these necessary conditions must be satisfied if we are to have a case of personal identity.

This means that we can formulate the Brain criterion of personal identity as:

A Person $P_2$ at $t_2$ is identical with a person $P_1$ at $t_1$ if and only if there exists a brain $H$ such that,

(i) $H$ is the brain of $P_2$ at $t_2$,

(ii) $H$ is the brain of $P_1$ at $t_1$.

To emphasise the intuitive plausibility of the Brain criterion as the criterion for personal identity over time, consider what we would think about a brain-transfer case. Suppose there are two persons, Brown and Robinson, who are both undergoing treatment for brain-tumours. Suppose ordinary treatment of these tumours has had no effect at all upon the health of either Brown nor Robinson. As a desperate way out of this situation, the surgeon suggests to Brown and Robinson that they should undergo an operation where the brains of Brown and Robinson are taken out of their skulls, treated with a new advanced and strong form of chemotherapy, and then put back into the skulls again. Suppose now that the surgeon tells Brown and Robinson that the treatment they are about to undergo is uncertain, since it is new, and that he cannot guarantee a successful result of the treatment. Brown and Robinson undergo this treatment, but, as it happens, the brain of Robinson is damaged during the chemotherapy and the surgeon, not keeping track of the brains, accidentally puts Brown's brain in the body of Robinson. Let us call the resultant conscious person with Brown's brain and Robinson's body 'Brownson'. What would Brownson think about the situation when he awakes? As Shoemaker says;
lously "That's me lying there!" (...) When asked his name he automatically replies "Brown." He recognises Brown's wife and family, and is able to describe in detail events in Brown's life, always describing them as events in his own life.\textsuperscript{119}

Is it not quite reasonable to suppose that Brownson in this case would be Brown, and that the reason that he is Brown is just that all those psychological features exhibited by Brownson are causally dependent upon the brain of Brown. That is, it is because Brownson has the same brain as Brown used to have that we can understand why it is the case that Brownson gives first-person accounts of the life of Brown, and because we can give this explanatory elucidation of the situation, it seems reasonable to uphold the view that Brownson is Brown.\textsuperscript{120}

\section*{6.3 Revised Brain criterion for personal identity}

Thus, according to the Brain criterion of personal identity a person is where his brain is. But, this theory of personal identity cannot be the whole story of what personal identity consists in. Firstly, it does not seem to be necessary that the whole brain must be preserved, but only a part of the brain. Several empirical findings indicate that persons actually have survived the loss of neural activities in large parts of their brains and even the actual removal of large parts of brain-tissue. Two examples are people who have suffered from cerebral haemorrhage and people with brain-tumours. Furthermore, we only seem to be using a small part of all possible connections available in the brain. This opens up the possibility of the plasticity of the brain. That is, if a certain part of the brain is disordered for some reason, the functioning of that part can be taken over by another part of the brain. For instance, since the centre of language is located in the left hemisphere persons who have suffered from cerebral haemorrhage in their left hemisphere usually lose their ability to use language. But, with some training it is usually the case that these persons can learn to use language by using other parts of the brain. These facts indicate that only a small part of a brain is neces-


\textsuperscript{120} This answer to the problem does not seem to be Sydney Shoemaker's answer to the same situation, since Shoemaker, in his book, claims that "it would be absurd to suggest that brain identity is our criterion of personal identity". The reason for Shoemaker to take this attitude is that we can imagine Brownson acting and talking just like Robinson, even though he has Brown's brain. But, this argument against identifying the criterion for personal identity as the brain criterion just misses the point that it is \textit{because we can, by referring to the continuity of the brain, give an explanatory description of why Brownson must awake with the first-person experiences of Brown, and not of Robinson, that the brain criterion is correct.}
sary for personal identity. This is also supported by a thought experiment by H. W. Noonan:

Let us suppose that half of a man’s brain is destroyed and then the other remaining half transplanted into another body with consequent transference of memories, personality and character traits (...) 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.121

Secondly, following Peter Unger, we can distinguish between two different aspects of a person’s “dispositional psychology”; a person’s “core psychology and her distinctive psychology”.122 A person’s core psychology consists of all mental capacities, or psychological dispositions which are common to all normal persons, such as the capacity for conscious experience, the capacity to form for oneself certain simpler intentions and the capacity to reason.123 A person’s distinctive psychology, in contrast, consists of psychological states which the person shares with some other persons, but not with all of them, and psychological states which the person shares with no-one at all. An example of the first is the memory of having tasted a Stella Artois, while an example of the other is the person’s memory of being born at a certain date and time to a certain mother.124 What is important is that it is the preservation of the core psychology of the person that is important for the survival of the person. A person need not, to survive from time $t_1$ to time $t_2$, have any distinctive memories of what happened at $t_1$, the only requirement needed is that his core psychology, i.e. his consciousness, intentions and reasoning exercised at $t_2$ is realised in the same brain as the core psychology at $t_1$. The reason for this is that a person might suffer from amnesia so that he cannot remember one single event or intention of his earlier life, but still survive as the person who earlier witnessed certain events, had certain intentions and wishes.

Furthermore, it is not the active performance of higher cognitive capacities that matter in cases of personal identity, but it is the capacity to have higher cognitive capacities that are the fundamental condition for personal identity. For instance, imagine that a person $a$ is involved in a car-accident. As a result of the accident, $a$ becomes comatose. During the coma, $a$ does not have any conscious

---

122 Unger, *Identity, Consciousness and Value*, p 68.
activities, even though a has the same brain that he had before the accident. Suppose further that a recovers from the coma and once again involves himself in conscious activities. If this happened, I think we would be inclined to claim that the person who awakes from the comatose period is the same person as the one who had the car-accident, even though there is a period between these two events in which the brain seems to be nothing more than pure material tissue.

Having said that it is the core psychology that is the foundation of personal identity over time, and that it is the continuity of only a part of the brain that seems to be necessary for personal identity, we need to restate the Brain criterion as:

\[
A \text{ Person } P_2 \text{ at } t_2 \text{ is identical with a person } P_1 \text{ at } t_1 \text{ if and only if there exists a brain } H \text{ such that}
\]

(i) \( H \) is the brain of \( P_2 \) at \( t_2 \),

(ii) \( H \) is the brain of \( P_1 \) at \( t_1 \),

(iii) enough of \( H \) is preserved between \( t_1 \) and \( t_2 \),

(iv) there is no branching of \( H \) between \( t_1 \) and \( t_2 \).\(^{125}\)

6.4 Criticism of the Brain criterion

Is the Brain criterion really what personal identity consists in? Firstly, let us take a closer look at the clause (iii) of the different specifications of personal identity. What the clause says is that the spatio-temporal continuity of the brain must be a one-one relation, if we are to have a case of personal identity over time, must be a one-one relation. That is, a person \( P_2 \) at a particular time \( t_2 \) can only be identical with a person \( P^* \) at another time \( t^* \) if and only if at \( t_2 \) there exists only one person who is sufficiently spatio-temporally continuous with \( P^* \). That such a clause is a necessary part of the specification of the Brain criterion of personal identity is obvious, since if we did not include it in our specification we could very easily end up in a situation in which the transitivity of identity would be

\(^{125}\) This last requirement has been emphasised by several different philosophers. See for instance Derek Parfit, Reasons and Persons, third edition, (Oxford: Oxford University Press, 1987), p. 204, Unger, Identity, Consciousness and Value, p. 109, Mackie, ‘Personal Identity’ p. 203 and Sydney Shoemaker, ‘Wiggins on Identity’, Philosophical Review 79, 1970, pp. 541-543. The idea of this last clause is that we should rule out the possibility of there being more than one equally good candidate for being identical with the original person. This will be discussed more thoroughly when we consider possible criticism of the brain criterion.
violated. If, for instance, at $t_2$ we had two different persons $P_1$ and $P_2$, both spatio-temporally continuous with $P^*$ to the same degree, then $P_1$ satisfies the conditions for being identical with $P^*$, and $P_2$ satisfies the conditions for being identical with $P^*$, but $P_1$ would not be identical with $P_2$. But with the inclusion of clause (iii), it is argued, we need not worry about such possible cases of violation of the transitivity of identity. If we at $t_2$ had two different persons, both spatio-temporally continuous with $P^*$ to the same degree, then, because of clause (iii) neither of these persons would be the same as $P^*$. That it is equally appealing as necessary to include such a clause in the Physical Criterianist analysis's of personal identity is true, but, the question is whether the inclusion of such a clause is legitimate in the first place. I certainly do not think so.

Consider for instance the following successful double-transplantation. As I said before, a person can survive with only one hemisphere intact. Let us now suppose that the left hemisphere and the right hemisphere of a person are functionally isomorphic, so that both hemispheres can fulfill the same functions equally well. Suppose that we manage to transplant the left hemisphere of the person $a$ into the skull of $b$ and the right hemisphere of $a$ into the skull of $c$, and that both $b$ and $c$ awake after the operation with enough psychological capacities to be persons. If it would have been the case that we only transplanted one of the hemispheres, say the left hemisphere into the skull of $b$, then, according to the Brain criterion $b$ would have been identical with $a$. Furthermore, if we had conducted the other operation and transplanted the right hemisphere into the skull of $c$, then, according to the Brain criterion $c$ would have been identical with $a$. But, now, the Brain criterion suggests, since both $b$ and $c$ are equally good candidates for being identical with the original person neither one of them is identical with the original person. But this conclusion simply seems incorrect. How is it that what is considered in isolation by the Brain criterion to be two successful operations, is considered as a failure when they are combined. That is, "How can a double success, be a failure?"  

---

126 Derek Parfit, 'Personal Identity', in Philosophical Review, vol. 80, 1971, p. 5. It is important to notice that Parfit thinks this to show that it is the psychological relation between different persons that are what we really should put weight on, an not whether persons are identical or not. There is, according to Parfit no problem in principle with two different persons being psychologically connected with an earlier person, since psychological connectedness need not be transitive. It is only if we interpret the problem case as showing us something about identity that it causes problem, according to Parfit.
6.5 The intrinsicality of identity

Another, and more important problem with the Brain criterion is that the inclusion of a clause like (iii) violates the metaphilosophical principle that identity is *intrinsic*. By ‘identity’ being intrinsic is meant that whether *x* and *y* is identical or not should only depend upon the relationship between *x* and *y*, or facts obtaining between *x* and *y*. Particularly, the intrinsicality of identity rules out the possibility that the identity between *x* and *y* could, in certain situations, depend upon whether there exists another object *z* which is external to *x* and *y*. In no way can facts about an object other than *x* and *y* influence the identity between *x* and *y*. I would claim, since clause (iii) makes the identity between a person P₁ and another person P₂* depend upon the external fact of there existing a person P₂, and that the inclusion of such a clause is necessary for the Brain criterion, then the Brain criterion of identity just cannot be an adequate criterion for personal identity.

Some philosophers have claimed that we simply cannot rely upon the truth of an intuitive principle such as the intrinsicality of identity.¹²⁷ They have either claimed that identity can depend upon external factors, or that the question of intrinsicality or extrinsicality need never to come up in the first place. I will understand both of these positions as claiming that identity is an *extrinsic relation*.

Given the intuitive appeal of the intrinsicality of identity, what reason can someone have for claiming that identity is an extrinsic relation? I think the main reason is that it is possible to construe situations where, if only *a* and *b* exist in the situation, we would like to claim that *a* and *b* are identical, and situations where, if only *a* and *c* exist, we would like to claim that *a* and *c* are identical, but that it also is possible to construct a situation consisting of *a*, *b* and *c*, where *b* definitely is not identical with *c*. Such cases are *reduplication* cases.

Consider, for instance, the classical problem of the ship of Theseus. Suppose that we take Theseus’s ship and rebuild it piece by piece, so that at the end of the rebuilding we have gradually replaced every single plank, bolt and beam of the original ship. The original planks, bolts and beams of the ship have been burned in a celebration of Zeus. Calling the original ship ‘*a*’, and the completely rebuilt ship ‘*b*’, it seems that most of us would be inclined to claim that *b* is identical with *a*. This is, of course, also exactly what we could expect from the intrinsicality of identity. It is only *a* and *b* that stand in a certain relationship to each

---

other. But, consider now the following scenario. Instead of rebuilding the ship, Theseus determines to put it down into pieces and place them in a junkyard. After a short time these planks, bolts and beams are found by Diotima. Diotima soon realises that the things in the junkyard are pieces of a ship, and she starts to put them together again. As it happens, she puts every piece in the place where it originally was placed, so that she has a ship which differs in no way from Theseus’s original ship. Still calling the original ship ‘a’, but now calling Diotima’s ship ‘c’, I think most of us would be inclined to claim that c is identical with a. Still, only one object, c, stands in a particular relationship to the original ship a, so we still have an intrinsic relation between a and c. But, now, is it not obvious that we can have both of those scenarios taking place at the same time. In such a case, then, it seems that we have two different ships, existing at the same time, which could be identical to the original ship. But, from the Indiscernibility of Identicals we know that the ships b and c are not identical, since they have completely different histories. If b would be destroyed in a war against Persia, c could still be put in a dry dock in Athens. Furthermore, since it is quite obvious that a and c are different ships, they do occupy different places at the same time, treating both of them as identical with a would violate the transitivity of identity.

It is in contexts such as the ship of Theseus that the extrinsicality of identity might not seem as implausible as it first appeared to be. What this example shows us is that we seem to be able to use different ways of determining identity questions, and, unfortunately for us, these ways need not always coincide in their applications. In the first case where a is identified as b we use a very straightforward spatio-temporal criterion in which we trace an object under a certain description, such as being a ship. In the second case where a is identified with c we also use a spatio-temporal criterion, but this time without tracing the object under a certain description. In this case we trace the object by its compositional parts. According to a view which favours the extrinsicality of identity, both of these ways of determining questions of identity are legitimate. But, one way can be overridden if conditions satisfying both ways of determining identity is present at the same time.

Since claiming that the identity of an object with an earlier existing object can depend upon whether there exists any other candidate which also satisfies the criterion for identity with that earlier object, the view that identity is extrinsic can be called the ‘Best Candidate Theory’. According to this view, it is the best candidate for identity with the original object which is identical with the original object. According to the best candidate theory an object o₁ at time t₁ which is
continuous with an object \( o \) at \( t \) is also identical with \( o \) if, considering all possible objects \( o_1, \ldots, o_n \) at \( t_1 \) which are continuous with \( o \), \( o_j \) is the best candidate for being identical with \( o \).

One such best candidate theory is Robert Nozick's "closest continuer theory". The idea behind the closest continuer theory is that it is supposed to be a schema in which different dimensions of continuity can be measured, and once measured, help us decide which kind of continuity should have more weight in a particular case. As Nozick says, the closest continuer theory gives a necessary condition for identity, since "something at \( t_2 \) is not the same entity as \( x \) at \( t_1 \) if it is not \( x \)'s closest continuer". There are two different versions of this theory according to what kind of facts one should take into the account of continuity: the local version and the global version. Common to both versions is the idea that spatio-temporal continuity and qualitative identity are the measurements, or dimensions by which we should determine which of several candidates is the best candidate, and consequently, which one is identical with the original object.

The local version asserts that an object is identical with an earlier object if the later object has the highest degree of spatio-temporal continuity and qualitative continuity with the earlier object. If no later object is sufficiently continuously close to the original object, then the original object has ceased to exist. Similarly, if more than one later object is sufficiently continuously close to the original object, but no later object is more close than any other object, then the original object has ceased to exist. Looking at our example about the ship of Theseus where we have two continuous objects \( b \) and \( c \) with \( a \), the local version of Nozick's closest continuer theory would favour the identity between \( a \) and \( b \).

The global version of the closest continuer theory asserts, on the other hand, that besides having the highest degree of spatio-temporal continuity and qualitative continuity, we also need to include a measurement which takes into account the longevity of the different candidates. Let us consider our case of the ship of Theseus again. Imagine that, from the local point of view, we determine that \( b \) is identical with \( a \), and that \( c \) is another ship. Imagine now that very shortly after all of this takes place that \( b \), for some reason vanishes, while \( c \) continues to exist. According to the global view, we should, then, claim that it is \( c \) which is identical with \( a \).

---

130 *Ibid.* p. 34
The problem with this global version of the closest continuer theory is that it is only intelligible in a four-dimensional ontology in which objects are extended over time in virtue of having momentary temporal parts in the same way as events and processes. I have already criticised the four-dimensional view of persons and I will therefore not go into a close criticism of the global version of the closest continuer theory from that perspective here. However, we can object that the global version, from the perspective of a three-dimensional framework, has the absurd consequence that objects can change their identity. According to the global view, we should, after finding out that $c$ had a much longer existence than $b$ claim that it is $c$ which is identical with $a$, and not $b$ as our first judgement. However, it seems to be strange that we ought to change our view of what object is identical with $a$ only because $b$ ceased to exist so much earlier than $c$. This is strange because the identity of an object, what the object is, does not change. No matter how much I dislike or like the fact that I am the object I am, I am stuck with being that object. Being a particular object is, in this sense, out of our control. Identity is, as we use the predicate, necessary. It is not the case that $a$ is identical with $b$ under certain conditions and not identical with $b$ under certain other conditions. If $a$ and $b$ are identical, they are so by necessity, i.e. under all circumstances. This means that the global version of the extrinsicality of identity is incorrect, since it is inconsistent with the "impossibility of conceiving of an entity’s not being identical with that which it is, in fact, identical".\footnote{David Wiggins, \textit{Sameness and Substance}, (Oxford: Basil Blackwell, 1980), p. 95.} Of course, it might be argued that $c$ was $a$ all along. It was only because we made a mistake that we thought $b$ to be identical with $a$. Now, this might be so, but, claiming this conceals the main idea behind the closest continuer theory; that identity is extrinsic.

But, in fact, there is another way in which an advocate of the extrinsicality of identity might rescue his theory from our objection. This needs a little bit of rephrasing of the original situation. We have, as a matter of fact, three different possible scenarios. The first scenario consists of the continuing replacement of matter of the original ship $a$ into the ship $b$. The second scenario consists of the continuing replacement of matter of the original ship $a$ into the ship $b'$, while also rebuilding a ship $c$ from the dismantled matter of the original ship. The third scenario is where we have a full dismantling of the matter of the original ship $a$, and where this is later put together as a ship $c'$.

In my objection against the extrinsicality of identity I took it for granted that ‘$a$’, ‘$b$’, ‘$b'$’, ‘$c$’ and ‘$c'$’ are all names of ships and that $c$ in scenario two is iden-
tical with $c'$ in scenario three. But, there is nothing in the description of the case of the ship of Theseus which forces an advocate of the extrinsicality of identity to claim that $c'$ is identical with $c$.\(^\text{133}\) Suppose that $a$, $b$, $b'$, $c$, and $c'$ really are ships. Now, suppose further, following Nathan U. Salmon, that there exist such things as hunks of matter constituting a ship.\(^\text{134}\) Given this, we have a certain hunk of matter which constitutes the ship $c$ and a hunk of matter constituting the ship $b$. Let us designate the hunk of matter which constitutes a ship with capitol letters in such a way that the hunk of matter $C$ constitutes the ship $c$. Now, given this, an advocate of the extrinsicality of identity might argue that it is the hunks of matter which stand as candidates in the story of the ship of Theseus. But, they do not stand as candidates for being identical with the ship of Theseus (since a hunk of matter cannot be identical with a ship), but as candidates for the relation $R$, i.e. "the hunk of matter which at a later time constitutes the ship of Theseus at that later time". The hunk of matter $C$ exists in the scenario three and in the scenario two. In scenario three $C$ is the hunk of matter which at a later time constitutes the ship of Theseus, since it is the only candidate for possessing $R$. But, possessing $R$ is not a necessary feature of $C$, since, ex hypothesis, $C$ is not the hunk of matter constituting the ship of Theseus in scenario two. This means that the ship $c$ does not exist in the third scenario, even though it is true that the hunk of matter $C$, which in scenario two constitutes $c$, exists in the third scenario. But this means that the ships $c$ and $c'$ are two different ships. What we have is the case that $c'$ is the same ship as $a$, but not the case that $c$ is the same ship as $a$.

The question is whether this solution to our objection is acceptable. As it seems, the advocate of the extrinsicality has saved his theory from abandoning the necessity of identity, through accepting that $c$ is not identical with $c'$. But, through not accepting that these two objects are identical, the advocate of extrinsicality must accept that the existence of a certain object at a certain time $t$ is dependent upon the existence or non-existence of another object at the same time $t$. He must, that is, interpret existence as extrinsical.\(^\text{135}\) But it is, to say the least, counter-intuitive that an object $a$'s existence depends upon the existence of another object $b$ which is in every respect causally unrelated to $a$.

\(^{133}\) This is typically the case when it comes to four-dimensional theories, since they claim that 'a', 'b', 'b''', 'c' and 'c'' are names of ships, but that there exists ship-stages, and that it is the ship-stages that are the candidates for "the ship-stage which at a later time constitutes the ship of Theseus at that later time".


\(^{135}\) Noonan, Personal Identity, p. 159.
This means, then, that the extrinsicality of identity can only be upheld if we are prepared to either (i) accept that an entity is not necessarily identical with the entity with which it is, as a matter of fact, identical, or (ii) accept that the existence of alternative candidates determines whether the thing exists or not. I cannot see any reason why we should accept either of these options. Both options are in conflict with the intuitive idea that we can trace an object through space and time without having to know what takes place in another object existing at the same time but at a different place. Furthermore, from the subjective point of view of the persons \( b \) and \( c \), it must be strange to find out that their survival depends upon what happens in a body distinct from their own. From \( b \)'s point of view, he would consider himself to be identical with \( a \), since \( b \) would seem to remember events and actions which \( a \) saw and did and continue to fulfil the intentions which \( a \) once had, and the same would apply to \( c \) from his subjective point of view. But, of course, from the point of view of \( b \) and \( c \), a decision which says that their first person experience of themselves as identical with \( A \) is wrong only because there exists another person who also has these first-person experiences must strike them as an arbitrary decision.\(^{136}\) It is, then, an intuitive metaphysical principle of identity that identity is an intrinsic relations and I conclude that the Brain criterion is not what personal identity consists in, since the Brain Criterion necessarily needs to include the extrinsic clause (iii) in its specification of what personal identity consists in.

6.6 The Body criterion and the Physical Realiser criterion

The other Physical Criterianist accounts of personal identity are the Body criterion and the Physical Realiser criterion. The Body criterion claims:

A Person \( P_2 \) at \( t_2 \) is identical with a person \( P_1 \) at \( t_1 \) if and only if there exists a Body \( B \) such that

(i) \( B \) is the body of \( P_2 \) at \( t_2 \),

(ii) \( B \) is the body of \( P_1 \) at \( t_1 \),

(iii) enough of \( B \) is preserved between \( t_1 \) and \( t_2 \),

(iv) there is no branching of \( B \) between \( t_1 \) and \( t_2 \).

The Physical Realiser criterion claims:

A Person $P_2$ at $t_2$ is identical with a person $P_1$ at $t_1$ if and only if there exists a physical realiser of a person's core psychological capacities $PC$ such that

(i) $PC$ is the physical realiser of the core psychology of $P_2$ at $t_2$,
(ii) $PC$ is the physical realiser of the core psychology of $P_1$ at $t_1$,
(iii) enough of $PC$ is preserved between $t_1$ and $t_2$,
(iv) there is no branching of $PC$ between $t_1$ and $t_2$.

The Physical Realiser criterion gets its support from the fact that the brain seems to be only one possible sustainer of higher level cognitive capacities. The brain is a plausible candidate for the criterion of identity over time only in virtue of the brain being the sustainer of higher cognitive capacities which are claimed to be necessary for an object to be a person. That is, it is the continuity of psychological features realised in or by the continuity of the brain that are fundamental for personal identity. In this sense, the continuity of the brain is only derivatively of fundamental importance, it is not, qua pure material object, important. Since it is the functional structure of the brain, and not the brain qua material object, that is of importance for the production of the capacity of higher level cognitive capacities, it seems that it is only a contingent fact that personal identity over time is grounded in the continuity of the brain. The brain, being the physical realiser of our core psychology, is only one of several possible physical realisers of core psychology, and consequently, the continuity of the brain is only one of several continuities that could sustain personal identity over time. This idea has been argued for by Unger. Imagine, for instance, that the work of my brain is taken over by structures that are not bunched together as in a typical animal organ [i.e. in the brain], but are spread quite evenly throughout the whole organism. In the less ambitious version of the case, as the nerve cells of my brain are killed, newly inserted nerve cells, much more widely dispersed, may take up their contributory role. In the more ambitious version, my central nerve cells are replaced not by any new nerve cells, or even by any organic matter at all, but by complex bionic circuitry that is strung throughout most of my body. In either version, rather than being placed near each other in my head, the new structures may be strung throughout my body so that, in late stages, my psychol-

\[137\] Unger, *Identity, Consciousness and Value*, pp. 140-141. Unger is stating his criterion in future-oriented terms, while my criterion is stated in a relation between a now existing person and an earlier person.

\[138\] Mackie, 'Personal Identity', p 201.


ogy is as much realized in my arms and legs as in my torso and buttocks, but in my head least of all. In either version, of course, the process may be both slow and gradual, occurring in small steps spread out over much lived time. After sufficient time, there will be, in my brain pan, only a dead brain that is, for the realization of psychology, completely unsuitable and ineffective.\textsuperscript{141}

The Body criterion, on the other hand, gets its main support from its intuitive plausibility. From a purely epistemological point of view, the reason that I can identify a certain person, for instance a friend of mine, in a large crowd, is because I know his bodily characteristics, and they enable me to individuate him from the other persons that are in the crowd. Though the Body criterion is used in an epistemological sense in this case, one might plausibly argue that what we mean by a term cannot be completely separated from the knowledge of how to use the term and the gaining of knowledge of when we are to apply the term.\textsuperscript{142}

To gain knowledge of under what circumstances it is correct to apply the term 'football-game' is to know, more or less, some of the distinctive features required of a football-game, such as the kicking of a round object, trying to score and trying to prevent opponents from scoring etc. In knowing circumstances such as these, we get to know when we can use the term 'football-game'. It is to be used when an open-ended list of circumstances such as the kicking of a round object, trying to score and trying to prevent someone other from scoring etc. obtains. Conversely, if we do not know when to appropriately use the term 'football-game', i.e., we do not know any circumstances under which we could apply the term 'football-game' nor any circumstances in which it would be incorrect to use the term, then it seems that we do not really know the meaning of 'football-game'. If we know these circumstances in such a way that we can correctly apply, or use the term under appropriate conditions, then, at least, we know part of the meaning of 'football-game'. What is true of terms such as 'football-game' is presumably also true of the term 'personal identity'. To acquire knowledge of, and then know under what circumstances we are prepared to use 'personal identity' is to know what we mean by personal identity. If it is something which I know about my friend, then it is what he looks like, i.e., I know his bodily characteristics. Certainly, I also know his mental characteristics, and sometimes I may even determine whether some encountered person is my friend by the mental content of this encountered person. But, knowing his mental characteristics is plausible, it seems, only because I can identify a certain physical body as the bearer of the mental states. It is the sameness, or rather, the

\textsuperscript{141} Unger, \textit{Identity, Consciousness and Value}. pp 107-108.

similarity of bodily characteristics between my friend as he is now with what he was before which constitute the circumstances under which I can get to know how to use 'personal identity' appropriately.

The plausibility of the Body criterion of personal identity over time is also obvious if we consider the fact that the ordinary way to determine the veridicality of the content of psychological states, such as memories, intentions and experiences, is to relate it to a physical entity. This is, of course, most obvious in cases of memories. Suppose, for instance, there are two persons, a and b, who both claim to be the murderer of z. a’s ground for claiming to be the murderer of z is that he has memories of firing a gun towards z and seeing z fall down on the floor. Let us call this claim of a, ‘C’ and the memory of a, ‘M’. b’s ground for claiming to be the murderer of z is that he has memories of firing a gun towards z and seeing z fall down in the street. Let us call this claim of b, ‘C*’ and the memory of b, ‘M*’. Now, the only way the police could determine whether a or b is the actual killer, is to test the veridicality of the memory claim of a and b. During the investigation to test the veridicality of a’s and b’s memories is to try to select information about the whereabouts of a and b for the time of the killing of z, since ‘memory’ is, as I said earlier, a causal notion. Suppose now, that the police found some evidence which established that a was, at the time of the killing, located in a town 700 miles from the place of the murder, while b was, at the time of the killing, in the same town, even the same street, as z. The only way they could come to know that fact about a and b is, of course, that they have done some kind of identification of a and b in virtue of an identification of a physical entity which was present at the time of the killing.

But, now, claiming that we can verify or falsify the memory claim of a by relating it to a certain brain does not seem to be particularly plausible. When a third person, (or the police), gets to know that C is untrue and that M therefore is a non-veridical memory, while C* is true and that M* therefore is a veridical memory he does not get to know this by knowing that C* and M* have a certain causal relation to a certain brain H which was at place p at time t, while M and C lack such a causal relation to H. I mean, how could he? He cannot observe their brains! Even though he might reasonably think that, if C is true, then M should have a certain causal relation to the brain H, and that if C* is true then M* should stand in a certain causal relation to H, he must infer the existence of a brain at the place p at the time t by first identifying another physical entity. This means that the reliance of the derivative importance of the brain for higher cognitive capacities rests upon a more primitive identification of a physical entity in which the brain of a person can be individuated. Now, this physical
entity by which we identify and individuate the brain is the body of a person, and consequently, the most intuitive and promising answer to the question of how we know that psychological states are veridical is that they are caused by your body. No matter how much we try to observe the intention N or the memory M of a person a, or the brain-state which causes the intention I or the memory M of a person a, we will never be able to directly perceive them. In this sense psychological states are forever concealed to us, and the knowledge of their existence will always depend upon knowledge of a certain body. That is, since it is reasonable to assume that a person continues to exist as long as it continues to be a mentally endowed physical entity and that the physical entity that we can perceive is the body of a person, it is also reasonable to assume that continuity of body is a necessary condition for personal identity.

6.7 The embodiment of persons

A further point in favour of the Body criterion and the Physical Realiser criterion is the way in which persons usually are embodied. I will not, at this stage, consider whether embodiment is a necessary condition for being a person. Though I do think persons are embodied in organised bodies, even biological bodies, the defence of that view has to wait until I discuss certain forms of Non-Criterianism. What I want to do now is to assume the Physical Criterianist claim that persons are necessarily embodied, and given that assumption, establish that a Physical Criterianist ought to understand persons as embodied in a physical body larger than the brain. I will first look at the M-predicates attributed to persons and what physical characteristics these seem to imply, and then I will consider the possibility of a person being a brain in a vat.

(i) Attribution of M-predicates: Even though scientific experiments are correct in giving an important weight to the possession of a persisting brain for the continuity of higher cognitive capacities, it is still questionable whether having a brain can be understood as a sufficient condition for being a person. Consider, for instance, our ordinary understanding of what a person is. Our paradigmatic understanding of a person is that of an object capable of actions and experiences. This is something which we strongly experience in our own cases, but also in cases where other persons are involved. Our ordinary understanding of persons as objects being capable of actions and experiences is shown in our ordinary use

---

of P-predicates and M-predicates. We use P-predicates to imply the possession, or at least the capacity for possessing, conscious activity of the person, while we use M-predicates to imply the possession, or at least the capacity for possessing, corporeal characteristics of the person.\(^{144}\)

The question is, now, given our ordinary conception of persons as agents of action and subjects of experience, what corporeal characteristics ought a Physical Criterianist claim that a person must possess?

It seems reasonable to think that a Physical Criterianist ought to maintain that it is the possession of a material body, and not the possession of a brain, which is the necessary condition for being a person. The reason for this is that the attribution of M-predicates to the person just does not make any sense if the brain is supposed to be the object of these predicates. M-predicates normally concern facts about bodies, not facts about brains. If the M-predicates only implied the possession of a brain, such M-predicates as ‘is red-haired’ and ‘weights 75 kg’ could never possibly be satisfied, since a brain cannot be red-haired, nor weigh 75 kg. But, interpreted as predicates which are satisfied by a certain person's body the M-predicates are meaningful, since the empirical facts which ground the applicability of the M-predicates to the person involve the existence of a head, eyes, legs, arms and a torso, and as such they seem to determine a causal relation, not between a brain and a person’s experiences, but between a body and the person's experiences and actions.\(^{145}\) Thus, from a paradigmatic point of view, it is in virtue of persons having bodies that we can apply different M-predicates to persons. Since M-predicates depend upon empirical facts about a body, and not upon a brain, this seems to favour a criterion of identity over time for persons in which the primary spatio-temporal physical entity is the body, and not the brain.

(ii) Brain in a vat: A possible objection to the idea that persons must be embodied in a body is that a defender of the brain criterion can convincingly argue that cases of brain-transplantations show us that having a functional brain is sufficient for being a person. Since intuitively we would claim that a successful brain-transplantation in which we put the brain of person P into a new body B would give us a person who would have changed bodies, this implies that having a body is not a necessary condition for identity. According to this view, having a body is important for being a person only because person’s brains typically are located in the skull of a body. But, being located in the skull of a body


\(^{145}\) Ibid. p. 92.
might be a contingent fact about us. It is not a necessary requirement that a functional brain is located in a body.

Consider the possibility of having a brain put in a vat which contains all relevant nutrients required for keeping a brain functioning. Let us suppose that scientists have found out a way to produce nutrients which can be directly assimilated by the brain, so that, in this case, we do not seem to need any bodily organs which assimilate and distribute the nourishment required for sustaining a living person. We can even imagine, just to make the case a little bit more stronger, that the brain which is in the vat, though it cannot perform any actions, in the ordinary sense of ‘action’, still has a very highly developed conscious life. As the brain is located in the vat it produces jokes, scientific theories, music and has emotions. This is not all, as the sciences have developed, scientists have found a way to construe a super-computer which is connected to the nerve-fibres of the brain, so that the subject of experience sustained by the brain in the vat has experiences of normal actions. If the subject of experience tries to raise his arm, for instance, then the super-computer will cause him to experience the raising of an arm. In this sense, the consciousness sustained by the brain in the vat, even though it cannot perform any bodily actions, still has experiences as if it performed bodily actions.

Though this is a very highly fictional example, a defender of the brain criterion would maintain that it shows us that having a brain is a sufficient condition for being a person, and that the having of a body is, at best, a contingent feature of human persons in the ordinary world. I will not argue against the idea that the “brain in the vat” is a logical possibility in the sense that a possible world could be inhabited by brains in a vat, and neither will I argue, as Putnam has done, that it is impossible that we, the persons who use earth-languages, are brains in vats. Instead, I will try to argue that it is doubtful that a person who is sustained only by a brain, i.e., like a brain in a vat, would be able to create a correct understanding and use of certain spatial concepts and indexicals. What I want to claim is that the ability to use spatial concept and certain indexicals, such as ‘here’ and ‘there’ is due to the bodily awareness of persons. This is most striking

See Hilary Putnam, *Reason, Truth and History*, (New York: Cambridge University Press, 1981), pp. 5-21, Putnam's argument for the impossibility of us being brains in a vat is that, since what we refer to by a term is a part of the meaning of the term, the claim that we are brains in a vat is self-refuting. If we were brains in a vat, the meaning of the term ‘brain in vat’ would not mean the same thing as what we mean by ‘brain in a vat’. When we refer to a brain in a vat, we mean the existence of a concrete brain floating in a liquid, but, if it was true that we really were brain floating in a liquid, what we would refer to is not a brain in a vat, but a physical state of a computer. Hence, Putnam concludes, the idea of us being brains in a vat is self-refuting.
in the ordinary case when we human persons acquire spatial concepts and indexicals. There are three different forms in which a normal person, i.e., a human person, is aware of himself as having a body: through an awareness of the body image, through a general awareness of spatial characteristics of his bodily parts, and through an awareness of spatial parts of his body.

(i) **Awareness of body image:** To be able to perform any actions in the world and to perceive external object of the world, we must have a point of departure, or a starting point in which the world is given to us as something which we are a part of, from which we do receive information. This point of departure is our egocentric frame of reference which is “immediately used by the subject in the direction of action”.\(^{147}\) This egocentric frame of reference is our system of primitive axes to which all other objects in the world are related. It is egocentric, since it consists of the natural axes of the body, and it is immediate, since it is based upon non-identificational awareness of these axes. The axes of the person constitute the ‘here-nes’ of the person, the informational foundation from where the person relates external objects as being “there”, “in front”, “behind”, “above” and “beside”. In relating objects to these axes, the person can perform actions towards these objects. For instance, having related a pipe as being in front of him, i.e. determined a certain external object as being at a certain relational location in his frame of reference, a person might stretch out one of his arms and grab the pipe. The egocentric frame of reference, since it consists of the natural axes of the body, does coincide with the body image, i.e., our awareness of our own body as being an object in the external world. So, in this sense, our awareness of our frame of reference is a way of being aware of where our body starts and ends.

(ii) **General awareness of spatial characteristics:** A person also has a general awareness of the spatial characteristics of his body, such as a sense of joint position and balance, kinaesthetic sensations and tactual perception. That this is something which we are aware of ordinarily is obvious if we consider the fact that many sentences about general characteristics of bodily states are immune to error through misidentification relative to the first-person pronoun. That some of the cases in which we make judgements about our bodily characteristics are immune to error through misidentification is clear from what we are inclined to say about the meaningfulness of the following judgements in which bodily properties are ascribed to an object; “Someone’s legs are crossed, but is it my legs that are crossed?”, “Someone’s head is pointing downwards, but is it my

head that is pointing downwards?’, and “Someone's legs are moving, but is it my legs that are moving?” Since it is senseless to ask the question of whom the knowledge is about in a case of immunity to error through misidentification, in which bodily ascriptions are attributed to an object, we must have a certain immediate awareness of our bodily parts, i.e., we must have a body image which enables us to directly know our general bodily characteristics.

(iii) Awareness of bodily parts: Persons also have a specific awareness of spatial parts of some body as being parts of a particular body which every person would be calling “my body”. In having a sensation, such as feeling a pain, we are aware of two different dimensions of the sensation; the first is the qualitative dimension, which consists of an awareness of the purely qualitative feeling of a sensation, and the second is the spatial dimension, which consists of an awareness of where this qualitative dimension is located. This second kind of awareness is a kind of bodily awareness, since it locates where, in what part of the body, a certain sensation takes place. That we have this kind of bodily awareness is obvious if we consider the fact that bodily sensations are accompanied by a certain direct, and non-inferential inclination to act in accordance with how the bodily sensation is epistemologically given to us in our bodily awareness. When a person feels a pain in his leg, he not only feels the painful experience, but also the location of the pain as being a pain-in-my-leg, and it is this spatial component of the painful experience which makes him act according to where the experience is felt to be located. This bodily awareness of sensations must be understood as an awareness of parts of a body, and not only as an awareness of a location in an egocentric space, since having the sensation of a pain in the foot presents itself just as a pain-in-the-foot. This is also true of phantom-limb sensations. Though there is no limb in which the sensation could occur, the experience of the sensation is as if it took place in a body part. In this sense, our awareness of our bodily sensations is located in specific body parts which fill up locations in our egocentric space. But, this kind of bodily awareness does not depend upon an earlier awareness of our egocentric frame of reference. The different kinds of bodily awareness which I have been presenting here must be understood as simultaneously present, each supplementing each other. As Brewer says, “the intrinsic spatiality of bodily awareness is sustained by its directly presenting certain part of the subject’s body as filling particular


149 Ibid. p. 299.
egocentric locations”,\textsuperscript{150} and “this presentation of the body as determinately extended in physical space depends in turn on the bodily sensation itself apparently being set in that body there”.\textsuperscript{151}

The question, now, is whether the capacity for bodily awareness is necessary for us learning how to use and understand the meaning of spatial notions and some indexicals? I think it is. It is, for instance, by having the egocentric frame of reference that a certain place can be assigned as being “here”. Relative to our egocentric frame of reference we can fix the spatial locations of external objects and act upon them from that point of departure. Fixing the spatial locations of objects in our referential system is to determine the use of spatial concepts and indexicals, since, by fixing the locations we also establish the conditions for applying a certain spatial concept or a certain indexical. For instance, it is only because the computer I am writing this sentence on is located at that particular place in relation to my referential axes that I am able to determine it to be in front of me, and not behind me. In this sense, our understanding of what is meant by spatial terms, such as ‘in front’, ‘behind’, ‘above’ and ‘beside’, and indexicals, such as ‘here’ and ‘there’ rely upon our egocentric axes.\textsuperscript{152}

Furthermore, if bodily awareness is a condition for having an understanding of certain spatial concepts and indexicals, then this implies that the brain in the vat, or the subject of the brain in the vat, would never be able to reach an understanding of ‘here’, ‘there’, and spatial notions such as ‘in front’, ‘behind’, ‘above’ and ‘beside’. The reason for this is that, contrary to what is the case in bodily awareness, the location of the brain, or the location of singular parts of the brain cannot be a part of the brain’s experiences; neither in the case where a sensation is located, nor as being the foundation of egocentric axes from which the person can perform actions. A subject’s basic frame of reference must be a body image in which the axes of the whole body of the person is part of the experience of the person. But this seems to imply that a person is embodied in a body and not in a brain, since having a body is a necessary condition for being able to learn and understand certain spatial notions and indexicals.

\textsuperscript{150} Brewer, ‘Bodily Awareness and the Self’, p. 300.

\textsuperscript{151} Ibid. p. 300.

\textsuperscript{152} Of course, we do not always need to use an egocentric frame of reference. Sometimes, as in determining that a certain computer is beside a certain book, we can take either the book or the computer as our frame of reference and relate the other item onto this basic reference-determining object. But, this kind of “object-centred” reference system is only possible if we first have determined the objects place through its relation to ourselves, i.e. to our primitive egocentric axes.
Though this seems to imply that a subject who has the ability to use spatial concepts and indexicals necessarily has a body, this does not follow from the premises. What it does establish, though, is that it is a necessary condition that a person is embodied in a body at some time during his existence. Acquiring some spatial notions and indexicals requires the existence of a body. However, this is completely consistent with the idea that a person can continue to exist without having a body once these notions are acquired. Though this is consistent with the argument, I still think we should try to resist the conclusion that it is sufficient for being a person that there is a functional brain. The reason for this is twofold:

Firstly, even though the example of the brain in the vat is construed as consisting of a subject of experience, i.e., as a person, who does not have a body, this example still relies, in a very fundamental way upon the idea of bodily embodiment of persons. Remember that the example was described as containing a subject of experience who experienced himself as having a body. This first-personal experience of having a body must have some weight in our evaluation of the thought experiment. Though not being the only foundation for our affirmative intuition, it still functions as the most important consequence-grounding feature of the described situation. It is the fact that the subject is aware of himself as embodied in a body that makes this hypothetical subject so similar to us, and if this feature is taken away from the content of the situation, we would not be equally inclined to claim that the brain in the vat is a person.

Secondly, even though it is not necessary that a person is an agent of action as well as the subject of experience, the idea of a person as an embodied agent and subject is the stereotype of what personhood consists in. Our notion of being a person is derived from our interactions with ordinary human persons who, paradigmatically, are embodied in bodies. True, we still count a mentally and physically paralysed human being as a person, but we do this, not on the ground of his having a brain which functions inappropriately, but on the ground that we extend our understanding of what it is to be a person from our ordinary conception, i.e., that of an embodied human person who has experiences and who performs observable actions. A paralysed human being is, in this sense, a person who has a malfunctioning body and mind. In this sense, our conception of persons is, paradigmatically, our ordinary notion of ourselves as being persons, and we relate, and extend our classification of what it is to be a person, not by strictly following some necessary and sufficient conditions for personhood, since a stereotype does not consist of necessary and sufficient conditions, but by
considering and evaluating an open-ended list of conditions which one object can satisfy more or less.\textsuperscript{153}

\textbf{6.8 Criticism of the Body criterion}

Even though the Body criterion and the Physical Realiser criterion are more plausible than the Brain criterion they cannot be understood as adequate criteria of personal identity. The reason for this is that they rest, as the Brain criterion did, upon a clause which establishes an \textit{ex post} condition upon the identity-grounding relation. Both the Body criterion and the Physical Realiser criterion maintain a clause which can be expressed as (iv) the spatio-temporal continuity of the physical object necessary for personal identity has not divided. I will only consider the case of the Body criterion, but the argument could be formulated \textit{mutatis mutandis} against the Physical Realiser criterion.

It has been claimed that the Body criterion is the only criterion compatible with the fact that identity is a one-one relation, because the Body criterion is immune to reduplication cases. That identity is a one-one relation means that it is a relation such that, necessarily, for every now existing person P* there is only one future person P# who can be identical with the now existing person, or, necessarily, for every now existing person P* there is only one earlier person P with whom P* can be identical. Other kinds of relations, such as \textit{have the same length as}, and \textit{being the brother of} need not be one-one, but can be one-many. It is clear that identity must be a one-one relation, since a one-many relation is incompatible with the transitivity of identity. It is only if identity is taken to be a one-one relation that we can preserve the condition that if A and B are identical, and A and C are identical, then B and C should also be identical.

The reason for claiming that the Body criterion is the only criteria compatible with identity being a one-one relation is that it provides the basis of distinguishing \textit{sameness} from \textit{exact similarity}, since "'same body' and 'exactly similar body' really do mark a difference".\textsuperscript{154} A tracing of the spatio-temporal continuity of the body of a person would, if carried out thoroughly, lead us to a

\textsuperscript{153} But, this kind of openness is not something vicious, since it does not imply vagueness. Vagueness is when we have a concept F which, once being applied to a certain object a, admits of a statement such as "a is a F to a certain degree (where degree can be more precisely specified). But, this is not the case when we extend our ordinary notion 'person'. An object which is determined to be a person is fully a person. In this sense, the openness of our stereotype of what it is to be a person admits of elasticity, not of vagueness.

\textsuperscript{154} Williams, ‘Personal identity and individuation’, p. 10.
complete account of the history of that person. Since we would inevitably have a case of reduplication revealed to us in such a case of tracing, we would be able to answer any identity question negatively if the tracing of the body reveals a one-many relation. In such a case, we would say that the resultant persons are not identical, but exactly similar.\textsuperscript{155} It is because the Body criterion, with its focus upon the spatio-temporal continuity of the body is verificational in principle that it can serve as a necessary and sufficient criterion for personal identity. We are supposed to trace the continuity of a body A and determine its identity with a body B only on the ground that there does not exist another body C which also is continuous with A. If there exists a body C, then A and B are different persons, while if there is only B, then we should conclude that A and B are identical. Williams would not say that both C and B are spatio-temporally continuous with A in such a case, since he thinks that the splitting of a body into two not only breaks down identity, but also continuity. As William says "the normal application of the concept of continuity is interfered with by the fact of fission, a fact which would itself be discovered by the verification procedure tied to the application of the concept".\textsuperscript{156}

This means, then, that Williams' argument for claiming that two resultant persons in a case of fission are exactly similar with the original individual rests upon the claim that neither one of them can be said to be spatio-temporally continuous with the original person. To this we need to reply two things: Firstly, Williams' claim that continuity "breaks down" in a case of fission just seems to be false. Consider for instance a case where one amoeba divides into two amoebae. There is no doubt that in such a case we consider the two products as being continuous with the original amoeba. There is nothing in the intrinsic relation between the two different amoebae and the original which would make it plausible to claim that they are not continuous with the original amoeba. The problem is whether it is possible to say that one of the products is identical with the original amoeba. And, this is not something we can know by only looking upon the continuity relation between amoebae, since the continuity relation is one-many while the identity relation should be one-one. The same situation obtains in a case where we suppose that a person can divide into two different bodies. If, for instance, a person P, as he is about to walk out through the door, divides into \( P_2 \) and \( P_2^* \) we should not hesitate to say that \( P_2 \) and \( P_2^* \) are spatio-temporally continuous with P. The intrinsic spatio-temporal relation between P and \( P_2 \), P

\textsuperscript{156} Ibid. p. 23.
and $P_2^*$ is the same regardless of $P_2$ and $P_2^*$ exist simultaneously or not. That is something which we would know. What we would not know, on the other hand, is whether $P_2$ or $P_2^*$ is identical with $P$. It might be the case that $P_2$ is identical with $P$, or that $P_2^*$ is identical with $P$, or it might even be the case that neither one of $P_2$ and $P_2^*$ is identical with $P$. But what we do know, though, is that if $P_2$ is identical with $P$, then $P_2^*$ is distinct from $P$, and conversely, if $P_2^*$ is identical with $P$, then $P_2$ is distinct from $P$. Also, we further know that, just as in the case of amoebae, the spatio-temporal continuity of the body of a person cannot help us in determining the question of identity of a person, since spatio-temporal continuity of a body is a one-many relation while personal identity should be a one-one relation.

Secondly, suppose for a moment that Williams is right in claiming that fission implies non-continuity between the original individual and the two resultant individuals. Even if this were true, the bodily criterion would still be false, since we would end up in a situation where personal identity over time depends upon the existence, or non-existence of another person, since the continuity of a body would depend upon the existence, or non-existence of another body. That is, to claim, as Williams has done, that $P_2$ and $P_2^*$ would, if both of them exist, only be exactly similar with $P$, while, for instance, $P_2$ would, if $P_2^*$ did not exist, be identical with $P$, is nothing else than claiming that who a person is at a particular moment depends upon whether there exists another person at that moment. But, of course, just as the Brain criterion cannot be correct if it needs to invoke a clause which makes personal identity an extrinsic relation, neither can the Body criterion nor the Physical Realiser criterion.

However, this is not enough. According to the bodily criterion, and the Physical Realiser criterion, it is a necessary and sufficient condition for personal identity that there exists a spatio-temporal continuity of a body of some sort. These criteria imply that an object, if it is to be a person, must have a body at every moment during its existence. But, even if we suppose that it is necessary that persons are embodied, we can ask ourselves whether embodiment is a sufficient condition for being a person. Is it not also the case that a body of a person is, by necessity, of a certain kind. The problem here, then, if it can be shown to be a problem for the Body criterion and the Physical Realiser criterion, is not so much that they are wrong in their assertion that persons must be embodied, but that the terms ‘body’ and ‘physical realiser of core psychological capacities’ are left completely unqualified. True, they assert that the continuing physical object should be the body of a living person, but this does not seem to help us very much. At the moment, from the information we have from the statement of the
Body criterion and the Physical Realiser criterion, we simply do not know what a living person is. These criteria are, at face value, compatible with the view that a table, a stone, a dolphin and a bionic entity can be the body of a person, as well as a human being, since we do not know what sorts of “bodies” persons have.
7. ANIMALISM

7.1 Introduction

In this chapter I will discuss the central ideas of Animalism, i.e., the ideas that (i) we, the individuals who presently formulate the question of personal identity, are animals of the species Homo Sapiens rather than persons, and (ii) that our identity over time consists in the identity of the animals that we are. This account is a natural extension of the physical account of personal identity in that it tries to establish in a more determinate way what kind of physical body we have. Animalism has been endorsed by Paul Snowdon, Michael Ayers, Eric Olson, and Peter van Inwagen.¹ I will present two different versions of Animalism: one which understands the terms ‘person’ and ‘human being’ as equivalent, and one which understands ‘person’ as a phased sortal term applicable to objects falling under the substance sortal term ‘human being’, amongst others.

I will claim that Animalism, though determining a plausible necessary condition for our identity over time, cannot be understood as an adequate criterion for our identity, since (i) it already employs a pre-understanding of personal identity over time, and (ii) that it relies upon a misrepresentation of which substance sortal it is that picks out what we are essentially.

7.2 General ideas of Animalism

In illustrating the Animalist position concerning personal identity, I think we should begin by considering certain facts concerning ourselves which we do find obvious. First of all, it is an acknowledged fact, a point which does not even

seem to be in need of mentioning, that individuals like you and I are persons. One of the most important elucidations of the notion ‘person’ is, as we saw earlier, Locke’s definition of a person as being

a thinking intelligent being that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places (...).\(^2\)

That you and I do satisfy this Lockean definition is not something questionable.

But, at the same time that we are persons, we are also members of a certain species, *Homo Sapiens*. That is, it is as equally undeniable that you and I are animals, as it is that we are persons. While you are reading this sentence your heart is pumping blood around your body, your brain is sending out and receiving information from different parts of your body, the digestive system of your body is constantly in action, as well as your respiratory system. And, ordinarily conceived, all of these functions of the biological organism, or the animal, are functions that you have, though you are unaware of some of them. That is, not only do we belong to the sort person, we are also animals (of some sort).

A central idea of Animalism falls back on David Wiggins’ characterisation of what is required for a correct determination of what it is to be a person.\(^3\) In his “animal attribute view”\(^4\) of personhood, Wiggins claims that one must determine what a person is through (i) relating it to a natural kind in which the real essence of persons is determined, or where it is stated what persons fundamentally are, and (ii) determine what a person is through a functional, or systemic, determination in which certain functions, or capacities, of persons are determined.\(^5\)

According to Wiggins’ animal attribute view

\[ x \text{ is a person if and only if } x \text{ is an animal falling under the extension of a kind whose typical members perceive, feel, remember, imagine, desire, make projects, move themselves at will, speak, carry out projects, acquire a character as they age, are happy or miserable, are susceptible to concern for members of their own or like species... [note carefully these and subsequent dots], conceive of themselves as perceiving, feeling, remembering, imagining, desiring, making projects, speaking ..., have, and conceive of themselves as having, a past accessible in experience-memory and a future accessible in intention ..., etc.}\(^6\)

Some Animalists might not accept the later part of this specification of personhood on the ground that whether an animal has certain psychological capaci-

---


ties or not is irrelevant to the question of what he is fundamentally. But, what all Animalists do accept is the animal attribute view, in so far as it is a determination of what kind ordinary persons, i.e., we, belong to. Included is the fact that we, the individuals of the species *Homo Sapiens*, function as paradigmatic instances of what is meant by 'person'. According to Animalism, though it is true that you and I are persons, individuating us as being persons is not the fundamental determination of what kind of objects we are. While it is true that we are persons, it is essentially true of us that we are animals. It is this latter sort which gives us the fundamental answer to what kind of objects individuals like you and I are. The fundamental substance sortal term which answers the *what is it*-question concerning us is ‘animal (of the sort *Homo Sapiens*)’, not 'person'.

Claiming that persons, as ordinarily conceived, are animals requires that we can say something about what it is to be an animal. A reasonable, and for our purposes sufficient, characterisation of animal is the following *animal condition*: an animal is a *self-sustaining biological unit in that* (i) *it is a complex entity who has evolved in an evolutionary chain and* (ii) *that it sustains its unity as a complex living object in virtue of keeping itself in homeostasis through certain biological functions such as metabolism, teleology and organised complexity.*

This characterisation of animal is sufficient, since it captures the central idea that animals have "life-supporting processes" which enable them to be structured units who can sustain their own unity by self-organised regulation. Only if an animal has a metabolism can it receive the necessary nutrients for retaining its "dynamic stability", or fundamental unity and structure. In this sense, as Eric Olson points out, an animal, or a living organism, resembles a flame that looses its "life" if not fed with new molecular raw material to consume. But, if an animal resembles a flame in that it needs new material to retain its dynamic stability, in another respect the animal is very different from a flame. A sufficient condition for a flame to retain its dynamic stability is that it is fed with new raw material from its external surroundings (fuel, oxygen, wind and temperature). However, when it comes to animals the intake of new raw material is teleologically determined by the internal structure of the animal itself. In this sense, animals are self-directed and self-organised objects that can sustain their own dynamic stability both in relation to their surroundings and in their internal structure. Different

---

parts of an organism, like those of a fine watch, are connected together in such a way that each has a role to play in enabling the organism to achieve its ends - survival and reproduction. No part can fulfil its function without the others, the entire structure will collapse - the organism will die and decay - unless all or nearly all of its parts do what they are supposed to do.  

Furthermore, the goal-directness of animals is based in the complex structure of the animal. First of all, animals are composed of an enormous number of cells which, if organised in another way would not constitute a living organism. But, secondly, this organised complexity of cells carries with it a sort of internal genetic plan, the DNA, which provides the animal with the basic operating instructions needed for repair and replacement of tissue, for the organism's growth and development, and for the reproduction of new animals of the same kind.  

Given this characterisation of what an animal is, then, one of the central claims of Animalism is that we are living organisms who have

"life-giving" features - metabolism, teleology, organised complexity - and whatever further properties necessarily go along with them, such as self-directed growth and development, an internal genetic plan, low internal entropy, and perhaps the capacity for evolution by natural selection.  

Animalists are also committed to accept the further claim that being the same animal is necessary and sufficient for the survival of individuals like you and me. That is, beside relying upon the animal attribute view in claiming that every individual like you and I are identical with an animal, an Animalist must also claim that the persistence condition for individuals like you and I is that of some kind of animal. In this sense, Animalism upholds the view that it is possible to specify a criterion of identity for individuals like you and I in terms of sameness of animal (of the sort Homo Sapiens). As Paul Snowdon once formulated it, Animalism is the view that

we are identical with a certain animal. But it is also part of the view that our persistence conditions are those of animals, animals being regarded as one fundamental kind of thing. So the second claim, which I can treat as at least partly elucidatory of this thought, is that anything which is an animal must be an animal, and the self-same animal, at all the times it exists.  

---

10 Olson, The Human Animal, pp. 128.
11 Ibid. pp. 128. Olson claims that a human animal, for instance, consist of $10^{14}$ cells containing $10^{12}$ bits of information.
12 Ibid. p. 129.
13 Ibid. p. 130.
14 Snowdon, ‘Personal Identity and Brain Transplants’, p.111.
That is, according to Animalism, not only is it a necessary feature of me that I am an animal, it is also impossible that I persist over time without continuing to be that animal, because the persistence conditions by which I survive, I do have in virtue of being an animal. This means then, that Animalism entails the following criterion of identity for individuals like you and I;

(i) *a person like you or I existing at a time \( t \) is identical with an animal (human being) existing at time \( t \),*

(ii) *an individual like you or I cannot cease to be an animal (human being) without ceasing to exist,*

and

(iii) *if \( x \) and \( y \) are individuals like you or I existing at \( t \) and \( t' \), respectively, then \( x \) at \( t \) is identical with \( y \) at \( t' \) if and only if \( x \) at \( t \) is the same animal (is the same human being) as \( y \) at \( t' \).*

The main purpose of these clauses of Animalism are that clause (i) is intended to establish that human persons are, fundamentally speaking, animals. Clause (ii) is not meant to imply that an individual like you and I cannot cease to be persons. What is meant is that clause (i) together with clause (ii) show that individuals like you and I survive over time in virtue of being animals, not in virtue of being persons. Clause (iii) is a specification of the criterion of identity for individuals like you and I, and it is that clause (iii) together with clause (ii) that makes Animalism a form of Criterianism.

### 7.3 Different versions of Animalism

Animalism can interpret the general approach to the problem of personal identity over time in different ways. Let me give an example by Paul Snowdon: you and I are in the desert and see in the distance an object. It is very hard to make out, but someone hazards the suggestion that it is a person. How would we respond to the remark? I think that there is an inclination to say that we would agree if, but only if, the object turned out to be a human being. We would not say, except as a joke, ‘You are quite right, because I can now see using these binoculars, it is a highly intelligent Martian’. Nor would we say, again except as a joke, ‘You are quite wrong, for, as I can now see using my binoculars, it is in fact a brain damaged man who has lost his powers of reason’.

---

15 Olson, *The Human Animal*, p. 18.
What this example by Snowdon shows us is, I think, the fact that there is a strong inclination to interpret 'person' as being equivalent with 'human being'. That is, the ordinary concept of *person*, i.e. the one we use in ordinary discourse, seems to be nothing but the same as the concepts *human being* and *man*. But, now, given that it is possible to interpret 'person' and 'human being' as interchangeable terms, it is possible for an Animalist to uphold that *person* is a substance sortal, a sortal concept which picks out what an object is fundamentally, or essentially. According to this form of Animalism, not only is it the case that it is impossible that I can persist over time without continuing to be an animal of a certain kind, it is also the case that it is impossible for me to persist without being a person. This version of Animalism accepts the animal attribute theory of personhood put forward by Wiggins. According to this form of Animalism, it is the case that the following two statements are true: (i) I am fundamentally, or essentially a person, and (ii) I am fundamentally, or essentially an animal (of the kind *Homo Sapiens*).\(^{17}\) Combining this view with the idea that objects belonging to a certain kind have definite persistence conditions in virtue of belonging to this kind, this form of Animalism also upholds the claim that (iii) persons and animals of the kind *Homo Sapiens* have the same persistence conditions.

This is a possible view that an Animalist can take, and it is quite a plausible view if one is prepared to restrict personhood to human beings. But I do not think that one should understand 'person' and 'human being' as equivalent terms as the first version of Animalism does. The reason for this is that it does not seem to be the case, or at least an argument has to be provided in favour for the view, that only human beings can be persons. We have a long tradition of thinking that entities other than human beings are persons. For instance, many people believe in the existence of God and angels, and according to these believers, God and the angels are persons. Others, following a tradition of science fiction and computer science see no problem in the claim that sophisticated robots or computers could be persons. Less far fetched examples are biological organisms of different kinds, such as dolphins, and certain higher primates. These kinds of animals, and maybe others too, have several of the requirements normally associated with personhood. They have a physical body, they seem to possess some basic kind of language or communication device, a social structure, a certain kind of self-image, and they certainly do resemble human beings in their internal structure, since they are biological organisms. If it is the case that these objects are persons, which we cannot rule out *a priori*, then it simply

---

\(^{17}\) Snowdon, ‘Persons and Personal Identity’, p. 34.
cannot be true that ‘person’ and ‘human being’ are equivalent. If there are other kinds than human beings that are persons, then the concept person will contain a larger group of objects than the concept human being. That is, in that case human beings will only form a sub-class of persons.

Partly because of the possibility of non-human persons, most Animalists have taken the view that we should repudiate the question of personal identity over time in favour of a question about our, identity over time. According to this form of Animalism, while it is true that I am a person and that I am an animal (of the kind Homo Sapiens) it is not true that I am essentially a person. It is, on the other hand, true that I am essentially an animal (of the kind Homo Sapiens). If we start our inquiry by asking what our, i.e. your, mine and every other human beings, persistence conditions consist in, and not by asking the question of personal identity over time, we will realise that person and human animal have distinct persistence conditions associated with them. Firstly, something might be the same human animal without being the same person. Secondly, since it might be the case that other biological organisms are persons, something might be a person without being a human animal. Let us only focus upon the case where we have the same human animal without having the same person.

Let us suppose, with Eric Olson, that you, as a result of a terrible accident, have a temporary heart failure. Because of the heart failure your brain does not receive any oxygen for ten minutes. As a consequence of the lack of oxygen all your neurons in the cerebral cortex have died of anoxia. Now, neurology and neuro-psychology have taught us that thought and consciousness require a functional cerebral cortex, so a consequence of the lack of neurons in the cortex is that your higher mental capacities are irretrievably lost. That is, you will never again be able to have conscious thoughts, or remember your past, since having a functional cerebral cortex is a necessary condition for being able to do these things. Now, while the cerebral cortex is absolutely destroyed, it is possible that certain subcortical parts of the brain, for instance, the thalamus, basal ganglia, brainstem and the cerebellum, continue to fulfil their functions. These lower part of the brain sustain respiration, metabolism, circulation and digestion, and these function can be fulfilled many years after the irretrievable loss of higher mental functions. Now, let us call the state that you have elapsed into a “vegetative state”. Now, the important question is whether you survive this horrible process, and if you do, what kind of entity it is that exists in the vegetative state?

We can all agree that before the accident happened to you, there existed a person and a human animal at the same place at the same time. That is, where you were there was a person and a human animal, since you, before the accident,
were the person and the animal. After the accident it seems reasonable to claim that we have, at least, a human animal that elapsed into a vegetative state. It is still alive, since all lower functions are working. What it lacks are higher cognitive functions, such as thoughts and consciousness. That is, in the vegetative state we seem to be stuck with a human animal without mental functions. Now, since we earlier claimed that you are numerically identical with a human animal, it does seem to be reasonable to conclude that you are the animal in the vegetative state. The only alternative to this claim is to maintain that you, at some point in the process of mental depriving, cease to exist and are replaced by a numerically distinct animal. But this idea just seems absurd. Partly because the subcortical functions are continuous between you before the accident and the human animal of the vegetative state, but also because the human animal in the vegetative state will have all of your physical characteristics. For instance, the fact that you have a scar beside your right eye due to a hockey-stick explains the fact that the human animal in the vegetative state also has such a scar. Furthermore, the human animal of the vegetative state will receive attention and care by your loved ones. All of this strongly indicates that it is you who are the human animal, and that you have lost all of your higher mental capacities. But, Olson now maintains, claiming that you, as you are in the vegetative state, are a person is not equally straightforward. The reason for this is that you no longer satisfy the capacity specification clause of the animal attribute view, since you do not perceive, remember, etc. The individual existing in the vegetative state has no features which could distinguish it as being a person from being a non-person.

A second example indicating that 'person' and 'human being' are different from each other is the fact that we do not seem to count a human foetus as a person, though we do count it as a human being. This is the "foetus case". Most people would agree that a human foetus is a human animal. We also know that this human foetus will, if circumstances are good, grow into an adult human person. The adult is a human being and a person. We also know that the human foetuses do not have a functional cerebral cortex until the twenty-fifth up to the thirty-second week after fertilisation. Before this crucial period in which the cerebral cortex starts functioning, it is at least puzzling whether we should call the foetus a person. It has no distinguishing psychological features of persons. But at the same time, there is no doubt that there is an individual in the womb, and that this individual is a human being.

---

18 Olson, The Human Animal, p. 17.
19 Ibid. p. 75.
What these examples show us, then, according to this form of Animalism, is that it is not the case that person and human animal have the same persistence conditions. The reason for them not to have the same persistence conditions is not that the person is distinct from the animal, but because the concept person does not determine any semantic-cum-metaphysically distinctive persistence conditions, since person is not a substance concept. Person is, according to this view, a phase sortal, since an individual need not cease to exist if it ceases to be a person.20

Earlier I referred to the distinction between substance concepts and phase sortals, and I claimed that a substance concept is applicable to an object at every moment that that very object exists, while a phase sortal picks out a certain limited phase of the object’s existence. I also said that a substance concept, contrary to phase sortals, usually gives semantic-cum-metaphysical persistence conditions for objects falling under the substance concept, in the sense that these persistence conditions should specify the coming into being of an object and the ceasing of being of that very same object. I also said that examples of such substance concepts are horse and frog, while examples of phase sortals are infant and baker. Now, the advocates of the second version of Animalism hold that when we ask a question of what I, you or we fundamentally belong to, we will see that being a person is something one can be at one time, but not be at another time. In this sense, the concept person is of the same category as infant and baker. I, for instance, am an adult who can carry out many complicated things, but at one time of my existence I was an infant who was unable to do even the simplest calculation or to walk. Similarly, today I am a philosopher trying to finish my work on personal identity over time, but it might be the case that I suddenly decide to quit being a philosopher and become a baker instead. In all of these cases of alteration, from infant to adult, from philosopher to baker, it is one and the same object that undergoes the alteration. Now, according to the advocates of this second form of Animalism, the same is true of the concept person. I, for instance, am now a person, but at one time of my existence I was a foetus that was not a person. It is one and the same object that was once a foetus and is now a person. Similarly, it might be the case that one day I stop being a person. For instance, if I loose all my higher cognitive functions but still remain alive in a vegetative state, that individual in the vegetative state, though not being a person, would still be me. Person, then, is a concept which picks out a certain phase of an object, and not what the object is fundamentally or essen-

tially. Since person is a phase sortal it is a mistake to associate it with a semantic-cum-metaphysical distinctive criterion of identity. Though it is true that we might ask ourselves what it takes for a person to persist through time, this is not to try to determine the coming into existence or ceasing to be of a particular object or individual. Doing this is on a par with asking what it takes for an infant or a baker to persist over time. It is, so to speak, not the determining of the existence of a substance, but only a feature of a substance. Since person does not determine any interesting semantic-cum-metaphysical persistence conditions, this means, from a semantic-cum-metaphysical point of view, that the concept person is of no interest to us. What is of importance is that we are human animals, and that we must remain human animals as long as we exist.

So the two striking feature of this version of Animalism is its denial that being a person is an essential property of a person, and that individuals like you and I are essentially human beings. Though it is true that you and I are persons, this is something which we are contingently. We might lose our personhood and still survive, but we cannot survive without being human animals. Furthermore, human persons are only one of several possible types of persons, and our, i.e., yours and mine, persistence conditions are persistence conditions that we have qua human beings. Other kind of persons might have other persistence conditions.

7.4 Two remarks on Animalism

Before evaluating this form of Animalism I want to clarify certain assumptions and consequences of Animalism. Firstly, it is an assumption of Animalism that it is not claimed that the criterion of identity for individuals like us must be conceptually, or a priori true. Consider, for instance, the case of a small child. A two year old child masters the linguistic use of terms such as ‘animal’, ‘dog’, and ‘horse’. From the behaviour of the child, if not also from the linguistic behaviour of the child, we have reason to think that the child masters the concept person. A two year old child refers to itself with the pronoun ‘I’, it plans for the future when it builds roads in the sand-pit, and it remembers earlier episodes of its life. Furthermore, it communicates with its parents and other children in a way that implies the mastering of the concept person. Imagine now the first time

---

21 Of course, the concept person might still be of interest when it comes to legal and moral issues, but that is, I reckon, something else than our deep interest in our identity over time.

22 Olson, *The Human Animal*, p. 27.
when this two year old child discovers the truth of the fact that he is a kind of animal. That is, he gets to know that he is a member of a biological species. What can we expect of such a situation? Well, some children probably would accept it immediately. But, and this is the important point; at least some children react as if they are taken by surprise. Some would even claim "I am not an animal". Now, the two latter responses are interesting because it indicates that defining 'person' and 'personal identity' in terms of animal and sameness of animal cannot be straightforwardly conceptual. If it was, a child who masters all the relevant concepts could not be surprised by the discovery that he is an animal. Furthermore, the denial of being an animal of a child who just heard that he is an animal is not a straightforward contradiction on behalf of the child. These facts indicate that claiming that every (human) person is numerically identical with a human being, and that the persistence conditions for that kind of being is the continuity of the life-supporting functions cannot, and should not be, understood as a straightforward conceptual analysis of the concepts person and personal identity over time. Instead, the Animalist analysis of what it is to be a person and what personal identity, or our identity, consists in should be understood as straightforwardly metaphysical. Since it is not a priori, it is an analysis which is to be determined by empirical investigations into what necessarily must be the case for the existence of individuals like you and me.

This claim of a posteriori components of the analysis of personal identity over time is absolutely consistent with the realist assumption of the subject of personal identity and my understanding of conceptual analysis: an investigation of the meaning of 'personal identity' which incorporates a posteriori components. To get the meaning of a term it is necessary to incorporate, as a part of the meaning, the fixation of the referent of the term to an actual object which functions as a paradigmatic instance of the term and the causal history of the term.

The last clarifying point we need to do concerning Animalism has to do with the interpretation of 'being the same animal as'. More precisely, is it the case that an animal, for being an animal, must be a living thing, or is it possible that there can be dead animals?

We have stated that Animalism maintains that an animal is a living biological organism, and that a human animal persists as long as the biological living organism that is the human animal continues to exist. That is, our way of preceding has determined that it is a necessary feature of human animals that they are living things, and that a human being ceases to exist when it is no longer a living thing. But this way of understanding Animalism faces two difficulties. The first has to do with the fact that we do speak of dead animals. For instance, we can
very often observe bodies of animals lying on the highway while driving in a car. The way we report this fact, when we see it, is that it is, for instance, a dead rabbit on the road. That is, we do not ordinarily say such things as: "I just saw the corpse, and the corpse was in the shape of what once was a rabbit". What we report is: "I just saw a dead rabbit".

The second difficulty with interpreting 'being a living thing' as a necessary condition of being a human being is that this implies that we cannot, strictly speaking, be identical with our biological bodies. A biological body remains intact, and to a certain extent continues to grow, after that the gross life supporting functions have stopped working forever. That is, the biological body has a history that extends the history of the human being (which must be alive to exist), and since different histories imply different individuals the human being cannot be the same as the biological body.

As I understand the issue at hand it is possible to resolve this difficulty for Animalism in either of two different ways. On the one hand an Animalist can maintain that it is impossible for a human being to exist without being biologically alive, and that calling something a dead human animal is only a figurative way of speaking. Strictly speaking, then, a human being is numerically identical with a functioning biological organism, i.e. a living biological organism, and not with the body, *qua* body. A consequence of this view is that there is no such thing as a dead animal. Once the life-supporting functions are gone, so is the animal. What remains is a corpse of what once used to be an animal. This reply is not necessarily inconsistent with ordinary thinking, since most of us would not mind having an autopsy carried out on what is, now, our bodies, after the biological life-supporting functions of the body have ceased to work. A consequence of this view, then, is that it is possible that we can have two different kinds of objects existing at the same time at the same place. In the case of human beings we have, beside the human being itself, also a physical, not to say, biological body.

The second way one can resolve the problem is to claim, with Michael Ayers, that we do not need to postulate that two different kinds of objects can occupy the same place at the same time. According to Ayers, a human being has life essentially in the sense that it is inconceivable that it (this human being) could have started existed as a non-living thing. The reason for this is that an individual human being is classified as a member of a natural kind through its origin

---

and structure. That is, it is because the individual human being is the offspring of a human couple and that it has the natural unity that it has that an individual object is a human being. But, now, it is reasonable to think that life is an inseparable function of its origin in the sense that if it was not the offspring of a human couple, it would not have been a living human being, and that life is the "principle of material unity" which can give an account of the structure and parts of the human being at every single time of its existence. But Ayers continues, it is important to notice that a living thing's current unity does not depend upon some present life-supporting processes, but upon previous life-supporting processes which determines the next moment of the living thing's existence. The same is also true when the life-supporting processes of a thing have disappeared. It is the life-supporting process of the human animal that determine the unity and structure of the dead human animal. If the life-supporting process had been different, or if it had ceased to function earlier or later in time, this would have meant that the history of the dead body would have altered too. In this sense, then, there exists a strong continuity between the living human animal and the dead body, and this continuity implies that it is one and the same thing that is alive and dead. That is when the individual human being dies, we should not understand this as the cessation of the object, or thing itself, we should understand this as a continuation of the human being as being a dead human being.

7.5 Criticism of Animalism

Let me now present a criticism of Animalism which shows that Animalism is inadequate as a criterion of what our identity consists in, even though I am sympathetic to the idea that being an animal is a necessary condition for being a person. The central ideas of Animalism are (i) that we are essentially members of the species Homo Sapiens, and (ii) that the question of our identity, qua what we are essentially, is more fundamental than the question of personal identity. These two ideas imply that Animalism firstly maintains the view that indexical terms, such as 'I', 'you' and 'we' refer, or at least fundamentally refer, to the human animal, and not to the person. In the foetus case, for instance, it seems clear that I am identical with a foetus even though the foetus is not a person. Secondly, it is also implied that Animalism entails that it is possible to deter-

mine what the identity over time of objects of our sort consists in without presupposing that we are persons. What I want to claim is that Animalism presupposes the concept person in determining that we are animals of the sort Homo Sapiens and that this implies (i) that they also presuppose a prior understanding of personal identity, and (ii) that the term ‘person’ must be understood as a substance sortal term.

According to Animalism, we are essentially animals of the kind Homo Sapiens. But ‘we’ cannot be understood in any other way than as short-hand for ‘individuals who are in their essential features similar to you and I’, since ‘we’ does not, in and by itself, delineate any determinate sort of objects. Hence, what the Animalist is really claiming is that individuals who are in all essential features like you and I are essentially of the kind Homo Sapiens.

But this means that whatever else ‘I’ refers to it definitely refers to a person. The reason for this is that the term ‘I’ cannot be understood in any other way than as referring to the individual determining the essential conditions of our identity. Firstly, the pure act of establishing that we are essentially human beings requires the capacity for experience, consciousness, reflection and self-conception. In establishing that we are human beings an Animalist must have a certain kind of awareness, or consciousness of himself as being a conscious and thinking individual which could be referred to by the first-person pronoun ‘I’. Even Animalists accept that these capacities are typically possessed by persons. In this sense, the very act of establishing that we are human beings is possible only if there is an individual who singles out, or individuates himself also as a person. Secondly, using ‘I’ as an individuative term is only possible if one has a prior individuation of oneself as being a person. Let me divide this point into four different steps:

(i) An individual who uses ‘I’ to refer to himself must be able to single out, or pick out himself as having experiences and possessing states of consciousness. Unless the speaker has individuated himself as having experiences, or possessing states of consciousness the use of ‘I’ in the referring expression would be in vain. Examples of such experiences and states of consciousness of the individual are mental features such as pains, emotions and thoughts. The speaker says such things as “I am feeling happy”, “I have a headache” and “I cannot really structure my thoughts concerning this argument”. All of this, then, shows that individuation of oneself as being a subject with mental characteristics is a necessary

condition for being able to refer to oneself with the use of ‘I’ in a referring expression.

(ii) Since the user of the first-person pronoun has individuated himself as being a subject who experiences things and who possess states of consciousness, it should be meaningful for him to ascribe certain P-predicates to himself, and for others to ascribe P-predicates to him.\(^{27}\) For instance, it should make sense for him to say such things as “I am in pain” and “I am depressed”.\(^{28}\) This means that an individual who is able to individuate, and refer, to himself, meaningfully can ascribe P-predicates to himself.\(^{29}\)

(iii) A necessary condition for the possibility of attributing P-predicates to oneself is that one also should be able to attribute the same predicates to other individuals. Not only should I be able to attribute ‘is in pain’ to myself, but I should also be able to ascribe ‘is in pain’ to you as well as to other individuals of the world. The reason for this is that P-predicates do not change their meaning when we attribute them to ourselves or to other individuals. For instance, if I say “I am in pain”, and “Peter Nilsson is in pain” I do not intend ‘pain’ to mean one thing in the first-person expression and another thing in second- and third-person expressions.

(iv) The only way in which we can attribute P-predicates to others is by accepting that the concept person is a primitive individuative concept by which other individuals behaviour is interpreted as the behaviour of consciously acting individuals. How would we be able to attribute P-predicates unless we have determined, in advance, that the individual is an object of which it is appropriate to attribute P-predicates. For instance, from observing individuals doing some body movements we cannot conclude that they have any mental life whatsoever. Besides the prior identification of the individual being an object to which P-


\(^{28}\) Of course, it might not be the case that the individual actually possesses the appropriate P-predicates for reporting what mental state he actually is in. It even seems reasonable to think that most of us experience states of consciousness for which we do not have an appropriate P-predicate. If we consider the many different pitches of pain, depression and happiness. Consider for instance headaches. We classify them all through a single P-predicate, ‘is having a headache’ but there are definitely different degrees of headaches, some being stronger and more painful than others. Some being like electric charges, while others are like buzzing drills.

\(^{29}\) This ascribing of P-predicates to oneself requires, of course, that the individual have the appropriate P-predicate in his linguistic repertoire. The question is not so much that the speaker must be able to formulate every single experience or state of consciousness in terms of a P-predicate, but that the individuation of oneself requires the mastering of at least some very basic P-predicates.
predicates are appropriate, there is nothing in the observational situation of an individual's movement of body which could force us to conclude that P-predicates are appropriate to ascribe to that individual. This means that we cannot, in and of itself, attribute P-predicates to the behaviour of an individual, unless we individuate the behaviour as being conducted by an experiencing and reflecting individual. But being an experiencing and reflective individual is, by all ordinary standards, to be a person. Consequently, the use of P-predicates in the first-person case relies upon the use of P-predicates in the second- and third-person cases. Since these latter cases of P-predicate ascriptions rely upon the concept person, this implies that the individuative use of 'I' is logically dependent upon person.

What remains now is to describe how it is possible, given the primitiveness of the concept person, to ascribe P-predicates to others. If we keep in mind that P-predicates do not have different meanings when ascribed to oneself and when ascribed to other persons, then learning the meaning of a P-predicate must involve, I think, two simultaneous processes. The first process contains the ascription of P-predicates to others on the basis of the observable behaviour of other persons, and the second process contains the self-ascription of P-predicates not based upon the observable behaviour of anyone. The first process gives us logically adequate criteria for applying the P-predicate, since it gives us, in principle, a way of determining whether the P-predicate is applicable to an individual or not. By 'logically adequate criteria' is meant that the observable behaviour is part of the very meaning of the P-predicate, not only a sign of the existence of a mental state of a subject. For instance, that a person cries, trembles and holds his hand against his cheek is not a sign of tooth-ache, but what we mean by 'tooth-ache'. The second process, though not being grounded upon any observation of behaviour, gives us the meaning of the P-predicate as the P-predicate figures as an element in a language. For instance, in ascribing 'Pär has tooth-ache' of an individual it is entailed that the individual Pär is having a subjective sensation of pain. But it is important to remember that these two processes function simultaneously. A learner of the meaning of a P-predicate must learn both aspects of its use, and the meaning of the P-predicate consists in both of these aspects.

30 Strawson, Individuals, p. 108.

31 Ibid. p. 110. Strawson gives us an analogy of a card. Though the markings of the card are logically adequate criteria for calling it, for instance the Queen of Hearts, these markings are not all that we ascribe to the card. We also ascribe to it, in the context of a certain game, properties over and above these marking. The same is also true of the logically adequate criteria for ascribing P-predicates. The behaviour of the individual is logically adequate
A consequence of the fact that *person* is the primitive concept by which a user of first-person pronouns must individuate himself is that Animalism involves a certain form of circularity, since it already presupposes *personal identity*. As I have said, individuating something as a person presupposes a three-dimensional ontology in which it is necessarily the case that persons who endure over time, i.e., for any two times that a person exists, that person is wholly present at those times. A way of characterising this fact is to claim that strict numerical identity obtains between a person as he is at a certain time \( t \) and a person as he is at another time \( t' \). I have claimed that this endurance view is part of the concept *person*, in that applying the concept *person* of an object entails that the object endures. If we combine this fact with the fact of the sortal dependency of individuation it follows that personal identity is the primitive level of description of our persistence over time. In individuating an object as being a person at a moment we also entail that that person is strictly identical with some person existing at another time. But claiming this is not different from claiming that personal identity obtains between an individuated person and a past or future person. This means that Animalists, since they presuppose the concept *person* in using first-person pronouns, cannot specify the criterion of identity for objects of our sort, i.e., individuals like you and I, without already employing an understanding of 'personal identity'.

Let me end this criticism of Animalism through providing a reason for thinking that 'person', contrary to Animalism, should be understood as a substance sortal term. I have claimed that certain sortal terms should be understood as phase sortal terms in that the terms refer to a contingent property of an object, while other sortal terms should be understood as substance sortal terms in that they refer to an essential property of the object. The three main reasons for upholding this distinction are (i) that it is conceivable that an object need not cease to exist even though a certain sortal term ‘t’ cease to be applicable to the object as long as another sortal term ‘t’ continues to be applicable of the object, (ii) that certain sortal terms are semantically complex in the sense that understanding the meaning of such a sortal term requires an understanding of another sortal term, but that the opposite need not be the case, and (iii) that semantically simple sortal terms are particularly suited for determining that an object is a *this such* in that they are applicable to the longest prolongation of the individual objects falling under the sortal.
That ‘person’ should, contrary to Animalism, be understood as a substance sortal term is due to a semantic property of proper names and indexicals. I have earlier claimed that the externalistic theory of meaning for natural kind terms is correct. I also accept an externalistic theory of meaning (the theory of direct reference) for proper names and indexicals. But, one must impose a certain kind of restriction upon the externalistic theory of meaning for proper names. Because objects are individuated via sortal concepts, it is a plausible assumption that there exists a semantically significant meaning associated with a proper names such that it is a necessary condition for a successful act of reference to an object \( a \) with a proper name ‘a’ that the user of ‘a’ associates, with ‘a’, the normally assumed sort \( S \) that \( a \) is individuated under.

To use an example of E. J. Lowe, suppose someone heard about the proper name ‘Aristotle’, but mistook the conversation to be about, say a book, and not a human being.\(^{32}\) Now, imagine that the individual is using the term ‘Aristotle’ in making the statement “Aristotle was great!”. Though it is true that Aristotle was a great philosopher, it is not true that Aristotle was or is a great book which is what the individual making the statement thinks is the case. But since the individual has mistaken ‘Aristotle’ to refer to a book, and not to a human person who was a great philosopher, the individual’s statement that “Aristotle was great!” is not, strictly speaking false or true, but out of place. It is, so to speak, a non-statement because there is no appropriate referent of the term ‘Aristotle’ and the correct way to approach the individual, when receiving more information about his wrong individuation of the term, is to say that he has completely misunderstood the meaning of the term. ‘Aristotle’, used as a name for an individual human person, rests upon a prior sortal individuation of Aristotle as being a person, and the subsequent use of the proper name must retain the sortal classification of the object. That is, if an individual grasps the use of a term, but in the act of grasping the term individuates the object referred to by the term \( \text{via} \) the wrong sortal concept, then that individual by his subsequent use of the term would not, strictly speaking, manage to refer to the ordinary object referred to by the term.\(^{33}\) This means that a proper name has a meaning, or sense, associated with it which determines “what sort of individual its referent is”\(^{34}\).

\(^{32}\) Lowe, *Kinds of Being*, p. 29.


\(^{34}\) *Ibid.* p. 29. But it is important to keep in mind that the kind of sense which I claim to be associated with proper names is different from the sense attributed to proper names in the traditional, or Fregean, theory of sense and denotation. The traditional doctrine of the sense of a proper name claimed that the sense should uniquely determine the referent of the term. But this is not what we have claimed. We only claim that the sort of the object referred to by
Starting out from this account of the semantic meaning of proper names I shall now argue that it is reasonable to understand ‘person’ as a substance sortal term. Imagine the following scenario. You have a grandmother called Samantha. When Samantha has used the name ‘Samantha’ and the indexical ‘I’ to refer to herself she has primarily used them as referring to what she has individuated as being the person that she is. During your younger years you individuated Samantha as a person and you learned that ‘Samantha’, or ‘Grandmother’ were names of a person. Furthermore, you learned that you were able to use ‘Samantha’ or ‘Grandmother’ to refer to the same object as Samantha herself did when using ‘Samantha’ or ‘Grandmother’. Imagine now that Samantha, by a terrible accident, enters into a vegetable state where all her former higher cognitive capacities are gone forever. Though she cannot think, she still has her biological functions intact. She breaths by herself and digests her food as before the accident. What I want to claim now is that the individual lying at the hospital is still your grandmother, Samantha, and that that individual is a person. It is a person who has lost all her higher cognitive capacities, but nevertheless, she is a person, since the individual lying at the hospital is still referred to by the name ‘Grandmother’ and ‘Samantha’. You say such things as “I’m going to visit Grandmother tomorrow”, “I wonder whether Grandmother is getting any better?” and “My Grandmother’s name is Samantha, and she is lying at such and such a hospital”. Other persons, for instance, the personnel at the hospital, speak about your grandmother as “Samantha in room twenty-four had low blood-pressure this morning!” and “Why don’t the grandchildren of Samantha visit her more often?” Now, there is no doubt that we do use language in this kind of way. But since a successful reference of the term ‘Samantha’ or ‘Grandmother’ in those circumstances to a particular object rely upon that object being a person, this means that ‘person’ should be understood as a substance sortal term. That is, it is essential for the term ‘person’ that being a person is an essential property of an object. In this sense, then, a person continuous to exist as long as he or she is a person. Ceasing to be a person is, for a person, the same as ceasing to exist. What all of this shows us is that we are essentially persons. We have to individuate ourselves as persons, and we have to be persons as long as we stay in existence.

the term should be semantically determined, not that the referent of the term is uniquely determined by the sense.
8. PSYCHOLOGICAL NON-CRITERIANISM

8.1 Introduction

The inadequacy and unintelligibility of Criterianism leave us in a position where we have to look for another solution to the problem of what our identity consists in. We have, then, to give a Non-Criterianist account of personal identity over time. I have claimed that the two necessary claims of a Non-Criterianist position of personal identity are the unspecifiability thesis and the primitiveness thesis. The first thesis denies the possibility of specifying personal identity in terms of some necessary and sufficient conditions, while the second thesis affirms the primitive nature of personal identity in relation to the empirical continuities that might underlie it. I also claimed that there are different Non-Criterianist accounts of personal identity and that they differ between each other on (i) the account given of why personal identity over time is unanalysable and primitive, and (ii) the account given of what this primitiveness of personal identity consists in.

In this chapter I will describe and criticise Psychological Non-Criterianist views, and particularly, I will describe and criticise Cartesianism, the Subjective View and Psychological Substantialism. What these views have in common is that they adopt a mainly first-person subjective perspective of personal identity implying certain assumptions about personhood. From this they conclude that persons are objects primarily of a psychological character.

8.2 Cartesianism

The two most striking features of the Cartesian understanding of personal identity over time, and hence what it is to be a person, are the ideas that (i) it is possible that a person can continue to exist even though the body of that person is annihilated, and (ii) what continues to exist in a case of personal identity over time is the soul.
The Cartesian idea that persons are distinct from their bodies and, hence, can survive without their bodies, I will call 'the possibility assumption'. This idea goes back to Descartes' claim that he can imagine himself existing without a body. The general arguments for this idea are (i) the non-essentiality of a person having the body that he actually has and (ii) that the disembodiment of persons is not a contradictory idea.

(i) Even if it would be true, which Cartesians deny, that every person must be embodied, it does not seem to be a necessary condition for a person to be embodied in one particular body. It is possible that you could have been given the body which actually is mine, while I could have been given the body which actually is yours. According to Cartesians, this shows that "none of the matter of which my body is presently made is essential to my being the person that I am" and, hence, that the person I am is distinct from the body through which I act, perceive, etc.

(ii) Furthermore, there is no doubt that we ordinarily think it is possible that a person may continue to exist without having a body closely attached to him. Consider for instance the idea of life after death. This is an idea which we very often meet in fairy-tales, literature and serious religions. This idea looks, from a logical point of view, self-consistent. We all understand what it would be like to continue to exist when our body has perished. It is the idea of having a subject of mental properties which continues to have these properties, while the body is destroyed. In all of these cases of disembodied existence what we imagine is a surviving person with consciousness, i.e., a person who has intentions, thoughts, feelings, fantasies and wishes, etc.

8.3 The person as soul

According to Cartesians, the thing that continues to persist in cases of disembodied existence is the soul of a person and therefore they conclude that persons are essentially souls. Souls, according to Cartesianism, are purely intellectual, or

---

spiritual substances whose only essential feature is Thinking, or Thought. It is a thing that “doubts, understand, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions”.

But a soul is not to be identified as the process of its activities. Using the methodological principle of non-identity, Cartesians claim that all of the activities of the soul are *modes*, or faculties of thinking which inhere in the soul. Cartesianism is, then, a version of substance-dualism in that it accepts that the world consists of two kinds of irreducible substances, matter and soul.

A consequence of this view is that ordinary human beings are complex objects consisting of both kinds of substances. The ordinary way to encounter a person is that of an object that has both a body and a mind. When I think of myself, for instance, I do think of myself both as being lazy, nervous, and as being 168 cm tall, weighting 55 kg. The same is true linguistically: I can truly utter such sentences as “I am lazy” and “I am 168 inches tall”. That is, ‘I’ can be correctly referred to myself both as having a mind, and having a body. On the face of it, using ‘I’ as a referring expression seems to refer to one object which is both a body and a mind. However, according to Cartesians, even though we use ‘I’ both in cases of self-attribution of mental properties and of self-attribution of physical properties this does not imply that we refer to one single object in both of these cases. According to the Cartesian, attribution of physical predicates to ourselves can be done only because we are contingently attached to a certain body which satisfies these predicates, while the attribution of mental predicates to ourselves is done in virtue of being attributed to our mind, or soul.

Swinburne has an argument which aims at establishing the truth of the existence of the soul. Two overall premises occur in the argument. Swinburne firstly accepts the possibility assumption. The second overall premise is what Swinburne calls the ‘Quasi-Aristotelian assumption’, i.e.

that a substance $S_2$ at $t_2$ is the same substance as an earlier substance $S_1$ at $t_1$ only if $S_2$ is made of some of the same stuff as $S_1$ (or stuff obtained therefrom by gradual replacement).

---

and that we allow that a substance might consists of “kinds of stuff other than matter”. What Swinburne means by this assumption is that it is incoherent to claim that \( x \) is identical with \( y \) if \( x \) and \( y \) have no parts in common. That is, if a substance is to continue in existence over time, then this continuity requires some continuity of the stuff that the substance consists of, and that this stuff can either be material or immaterial.

The argument by Swinburne is that from the fact that the Quasi-Aristotelian assumption implies that a later substance must have something in common, either material or immaterial, with an earlier existing substance if they are to be identical, and that it is logically possible that I may survive the death of my body, it follows that there must be some immaterial stuff that continues to persist in the survival of a person. But since the soul is defined as a non-material object, we can plausibly call this necessarily continuing immaterial stuff the soul. From the mere fact that I may survive the death of my body only in virtue of having a soul in common with the future person, it also follows that I, who exist now, must have a soul. That is, the fact that I exist now, implies that I now have a soul. In this sense a person is a combination of an essential part, the soul, and a contingent part, the body. The body and the soul form a unity which is affected by the surroundings, and the proper purpose of the body is “to inform the mind of what is beneficial or harmful for the composite”.

8.4 The Subjective View

Another Non-Criterianist version of personal identity has been advocated by Geoffrey Madell. According to this view, the direct subjective presentation of the person implies (i) that the identity of persons cannot be specified in any empirical conditions on the ground that our own persistence does not entail, or imply, the continuing existence of any empirical continuities, and (ii) that a person is nothing but the mental entity directly accessible in the subjective dimension of experiences. I will call this view ‘The Subjective View’.

According to ‘the Subjective View’, from a first-person point of view a person cannot identify himself with any empirical conditions, since determining or knowing that a certain description is true of himself presupposes a prior acquain-

---


167 Descartes, Meditations, pp. 56-57.

tance with the person that he is.\textsuperscript{169} For instance, to be able to determine whether the experience of seeing a computer is \textit{my} experience it is not enough for me to know that there is a person in the world at this moment who is seeing a computer. This knowledge, in and of itself, does not imply that the experience is \textit{mine}. To be able to identify the experience as \textit{my} experience I must already have a direct knowledge of myself as the person who is having the experience in question. In this sense

I can only know myself to be that one and the same person to whom these various descriptions apply if I have knowledge of myself, an awareness of myself, which is independent of these descriptions.\textsuperscript{170}

In this sense, ‘I’ is a logically proper name in that it is possible to know the referent of ‘I’ by direct acquaintance.\textsuperscript{171}

Since it is true that acquaintance with the self is prior to a description of the self it seems to follow that any description which we think is essential to the self really must be considered as only contingently attached to the self. More particularly, this implies that having a body or having experiences cannot be considered as necessary conditions for being a person at a particular time. In this sense, even though it is true that I have had a body or experiences at every singular moment I have existed, the having of this body or any experience are not an essential property of being \textit{me}. The I, the self, or the person is, in and of itself, something other than his body and co-conscious experiences, something which just cannot be contained in a description of physical and psychological terms.\textsuperscript{172}

What is true of a person at a time is also true of the person as he persists over time. Take the intuitive idea that it is possible to give a \textit{radical alternative biography} of a person’s life. For instance, it is certainly true that I, who actually am working in the field of philosophy, could have been a fisherman in the north of Sweden. This is an \textit{alternative biography} of my life in that it I imagine that the person who actually did a certain action \textit{A} did not do that action, but performed another action \textit{B}. But it also seems possible to think that I was a fisherman in the thirteenth century and in this sense I have radically altered the biography, since the biography no longer needs to be based in any prior actuality. Neither a certain physical origin, nor a mental origin seems to be necessary for being the per-

\textsuperscript{169} Ibid. p. 24.
\textsuperscript{170} Madell, \textit{The identity of the self}, p. 24.
\textsuperscript{171} Ibid. p. 24-26
\textsuperscript{172} Ibid. p. 25.
son that I am. The biography, though, is still a biography of my life according to
the Subjective View, since I can imagine myself to be that person in that there is
no difficulty in conceiving of the possible world in which the flow of experience
leading up to my present state of consciousness began earlier or later.\(^{173}\)

But, the fact that we can construe radically alternative biographies of persons’
lives implies that Criterianism cannot be correct in maintaining the view that
personal identity over time consists in the satisfaction of some physical and/or
psychological continuities which jointly are necessary and sufficient conditions
for personal identity. Personal identity over time must be equally unanalysable
as the first-person perspective of oneself at a particular time. Neither in the case
of present tense first-person thoughts, nor in the case of thoughts about personal
identity do we imply that the person, or the self, can be captured by a description
only containing objective, empirical continuities. Both cases show that “the
property of being mine is unanalysable”,\(^{174}\) and that it is
impossible to analyse what it is for a set of experiences over time to be the experi­
ces of the one person in terms of whatever objective relationships hold between
those experiences.\(^{175}\)

8.5 The subjective person

According to the Subjective View, what enables us to tell radically alternative
biographies of persons is that persons’ first-person perspective introduces an
“extra dimension of subjectivity”\(^{176}\) into the world. It is an extra dimension of
subjectivity in which every person is directly acquainted to himself and to which
the person would use ‘I’ or ‘self’ as referring expressions. The subjective sphere
of persons is an irreducible part of what is a complete description of the world.
But, then, what kind of entity is this person, or self, which cannot be analysed
and of whom each one of us is directly acquainted with in our own first-person
experiences?

As we have seen neither physical properties nor objective psychological prop­
erties are essential properties of persons. These kinds of properties are only con­tin­gently associated with the self as the self is at a particular time and over time.

---


\(^{175}\) *Ibid*. p. 25.

\(^{176}\) *Ibid*. p. 87
In this sense, a person need not be embodied. As we have seen, a person might have a radical alternative life history, so that the experiences that a person has during his life time are very different from what is the case in the actual world. The same is of course true if we consider the having of one experience at a particular time. For instance, instead of my having an experience of a computer at this very moment, it might have been the case that I had an experience of a stream of water. This means that the person, or the self is not his experiences. A person is something beyond the actual experiences that he has.

Since the reason for denying that persons are identical with their physical and/or psychological properties is the "extra dimension of subjectivity" of persons this implies that the Subjective View maintains that persons are essentially psychological beings. Though a person is not to be identified with the psychological features which it actually has, it still needs to have psychological properties. But a person is not to be identified with a Cartesian ego. It is an unanalysable entity which we are acquainted with in our first-person thinking, and where the experiences of the person, in and of themselves, are the "fundamental bearers of subjectivity". The experiences that a person, or a self, has consists of two different levels. On the one hand, we have the first-order level of an experience, which contains the element that we would be inclined to claim the experience is an experience of. For instance, my present experience in typing these sentences is, in this sense, of a computer. On the other hand, in having first-order experiences, we also have a second-order experience in which we directly experience the first-order experience as being our own experience. In this sense, every first-order experience is accompanied by a direct presentation of the experience in question as being mine. Since identifying the experience is to directly identify the experience as my experience, we just do not need to posit any mysterious entity sustaining our experiences. That entity is just the observable experiences themselves. As, for instance, Madell claims:

177 Madell, *The identity of the self*, p. 87

178 In this sense, the account of the self that the subjective view is giving resembles Kant's *transcendental ego* in lying as a formal accompaniment to the experiences a person actually has. But, according to the Subjective view, one should not identify the self of the subjective view with Kant's *transcendental ego*, nor with Descartes' *spiritual ego*. According to, for instance, Madell, both Kant and Descartes went further than their experiences gave them good reason to go. Both Kant and Descartes posited a mysterious a something we-know-not-what, to account for the identity over time of persons, or selves. But this is, according to Madell, to commit the same mistake as the Criterianists in their effort to make selves into objects.

the continuity of our experiences is intelligible in itself, it is the continuity of experience which allows us to talk about persons and not about a random sequence of experiences.\textsuperscript{180}

According to the Subjective View, the self as such “has no essential properties apart from ‘thinking’”,\textsuperscript{181} but this ‘thinking’ which is essential to persons is filled with experiences that present themselves in an irreducibly subjective way. In this sense, a person, or a self, is nothing but an undifferentiated sequence of experiences whose single basic property is the capacity for having experiences with an irreducible subjective component or dimension. A person, then, is essentially an unanalyzable psychological entity that, though still being distinct from its actual experiences, has first-person experiences in which the subjective point of view is given to the person.\textsuperscript{182}

8.6 Criticism of Cartesianism and the Subjective View

What I would like to do now is to show that, even though Cartesianism and the Subjective View is correct in claiming that personal identity over time is an unanalyzable and primitive fact of the world, they are wrong in their characterization of what it is to be a person. The reason for this is (i) that the approaches are based on three Cartesian mistakes, and (ii) that the first-person approach is methodologically wrong in that it omits certain fundamental truths about persons, such as that having a body is a necessary condition for being a person.

(i) I will divide the Cartesian mistakes into three different categories; (a) the fallacy argument, (b) the mistake from the ways of thinking of ourselves, and (c) the imaginability mistake.

(a) Both Cartesianism and the Subjective View wrongly assume that the first-person knowledge of ourselves is incompatible with the truth of us having certain essential properties that we are unaware of in knowing ourselves from the first-person perspective. According to these views, since (i) I have knowledge of myself, and (ii) this knowledge of myself cannot depend upon something of which I am unaware, and (iii) that I am unaware of everything except the referent of I, then this entails that the thing of which I am unaware cannot be essential to me being the person that I am.

\textsuperscript{180} Madell, \textit{The identity of the self}, p. 10.
\textsuperscript{181} \textit{Ibid}. p. 25.
\textsuperscript{182} \textit{Ibid}. p. 137.
But of course, the fact that I can have knowledge of a certain thing without thereby knowing its essential features is not sufficient to show that the object of which I have knowledge lacks certain essential features. For instance, just as it is true that I might have knowledge of water without knowing that water is essentially \( \text{H}_2\text{O} \) it might be the case that I have knowledge of myself without knowing that I am essentially \( \Phi \). But, from the fact that I can have knowledge of water without knowing it to be \( \text{H}_2\text{O} \) it does not follow that water is only contingently \( \text{H}_2\text{O} \), and nor does it follow from the fact that I have knowledge of myself independently of knowing myself to be \( \Phi \) that I am only contingently \( \Phi \).

(b) The second Cartesian mistake can be divided into two different versions. The first form consists of claiming that the fact that I can think of myself without thinking of myself as being \( \Phi \) implies that I am really not \( \Phi \). I will call this "the negation mistake". Contrasted with the negation mistake is "the positive mistake". The positive mistake consists in claiming that the fact that I can think of myself as having consciousness, or that thinking of myself is only possible if there exists some consciousness, implies that I am essentially a consciousness.

Just as it is possible for me to know myself without thereby knowing my true essential nature, I might also be able to think of myself without thinking of my true essential nature. For instance, from the fact that I can think of myself without thinking of myself as having a body or think of myself as having a different origin than that which I actually had it does not seem to follow that having a body is only a contingent feature of me nor that I or any other person could have had a different origin than the one we actually had. It might be the case that I am deluded due to the use of first person thoughts in thinking of myself without a particular body.

Furthermore, it might very well be the case that thinking, or consciousness, though being a necessary requirement for asking the question of what the nature or essence of a person is, only is a contingent property which certain persons have. For instance, just as we understand a sleeping unconscious person as a person, I think we would count an individual who has lost his consciousness due to an accident a person. Imagine, for instance, that your son or daughter had an accident. As a result of the accident the child has lost not only the presence of consciousness, but also the capacity for having consciousness due to a severe damage of the brain. Regardless of that we would treat the individual lying in the bed as a person. So, it seems that it might be the case that the capacity of thinking of oneself from the first-person perspective, i.e. consciousness, is only an accidental feature of persons.
(c) Both Cartesianism and the Subjective View are working with a 'possibility assumption' in which it is assumed that the conceivability of consciousness being distinct from the body of a person implies that it is possible that a person, or self, survives after the annihilation of the body, or that the conceivability of radically alternative biographies of persons implies that having a certain origin is not essential for persons.

But, from the mere fact that one can imagine, or conceive, the possibility of a certain state of affairs it does not seem to follow that that state of affairs is metaphysically possible. Consider what would be true about imaginability if it was true that persons had a soul, and what would be true about imaginability if persons are essentially physical entities. The first case, then, is the following; given that it is true that persons consist of a soul, then it is metaphysically impossible that persons could exist without a soul, i.e. there is no possible world in which a person exists and in which he does not consist of a soul. If this is true, then it is conceivable, i.e. logically possible, that persons have a soul. The second situation, given that it is true that persons must have physical parts, is that it is metaphysically impossible that persons could exist without physical parts, i.e., there is no possible world in which a person exists and in which he does not consist of physical parts. But, of course, if the latter is the case, then it is nevertheless conceivable, i.e., logically possible, that persons might have had souls. So, regardless of the metaphysical truth of souls or not, it is still imaginable that persons have souls. The same is true in cases of radically alternative biographies of persons. Given that it is true that origin is not essential for being the person one is, then it is metaphysically possible that a person could exist with a different origin, i.e., there is at least one possible world in which a person does exist and in which he does not have the origin which he in fact had. If this is true, then it is conceivable, i.e., logically possible, or imaginable that a person was born to different parents at a different time and place. Given that it is true that origin is essential for being a person, then it is metaphysically impossible that a person could exist with a different origin, i.e., there is no possible world in which a person exists and in which he does not have the origin which he in fact had. But, of course, if the latter is the case, then it is nevertheless conceivable, i.e., logically possible, or imaginable that a person might have been born to different parents at a different time and a different place. That is, we can think of this possibility in the sense that we can imagine the possibility. But, this imaginability does not show that a different origin is a metaphysical possibility. That is, imagining myself having a different origin does not imply that I could have had a different origin. What these examples show is that the arguments for the existence of the
possibility of disembodied persons or radically alternative biographies are arguments which show, at best, that we can imagine that persons have a soul or could have had another origin than what they actually had. Neither one of the arguments has established that we are, metaphysically speaking, distinct from our bodies. What it implies is that I can imagine that I am distinct from my body and that I was born at another time and place.

8.7 The necessity of physical embodiment of persons

(ii) Is there some argument which would settle the metaphysical issue of personhood? I do think there is such an argument, and this argument points in favour of understanding persons as material objects. The argument consists of two steps. Firstly, we have to acknowledge that Cartesianism and the Subjective View, in reaching their identification of the person with consciousness, have founded their views about persons on the wrong methodological foundation, and secondly, that the applicability of mental predicates to persons is ontologically dependent upon the applicability of material predicates to persons. As a consequence of the first step we have to conclude that persons are objects which are ontologically prior to consciousness, and as a consequence of the second step we must conclude that having a body is a necessary condition for being a person.

Firstly, in reaching their views that persons are essentially consciousness, Cartesianism and the Subjective View have taken an exclusively first-person perspective, or subjective point of view. Though I agree with Madell in the claim that experiences have the extra dimension of subjectivity, and that the first-person perspective presents the person as being nothing but the property of "self-hood", I think it is wrong to think it is possible to reflect upon the use of the notions ‘I’, or ‘the self’ and conclude that persons must be in a certain way. The reason for this is that ‘consciousness’ is not a conceptually primitive term, since ‘consciousness’ is not a fundamental term by which we can individuate an object. In this, Cartesianism and the Subjective View are susceptible to a similar

---

183 In what follows I will treat “consciousness” as a collective term standing for both Madell’s view of what a person is, and for the Cartesian notion of a soul. Even though there are differences between Madell’s and the Cartesian understanding of what kind of entity this consciousness is, I think it to be legitimate to do this on the ground that Cartesians identify the soul with the mental, or conscious part of the person. That is, for both Madell and the Cartesians, what is essential for being a person is the existence of consciousness, or the capacity for consciousness.

kind of criticism as Animalism when Animalism maintained that *person* is not the fundamental sortal by which we individuate ourselves.¹⁸⁵

‘Consciousness’ cannot be a sortal term, since it cannot enable us to single out different individuals in the required way. At best it can give us, if it is considered as a fundamental individuative term, the singling out of an individual, a consciousness, which is different from individuals of other sorts, but it cannot distinguish between individuals which are of the same sort. If ‘consciousness’, ‘I’, or ‘the self’ is understood as the primitive classificatory terms, then we would never be able to individuate different selves. That is, even if it is true that experiences are given with an extra dimension of subjectivity, so that every experience is accompanied by the property of being *my* experience, this account of experiences cannot ground a distinction between an experience being mine and an experience being someone else’s. The extra dimension of subjectivity, since it is a part of every single experience I have, would imply that every experience that exists is *my* experience.

But, as we all know, we can and do distinguish between experiences being mine and experiences being someone else’s, since we understand that there are other persons in the world who also refer to themselves with the first-person pronoun. This means that the individuative use of ‘I’ is logically dependent upon some other fundamental classificatory term by which we individuate individuals who have consciousness. I have already claimed that P-predicates are attributable to oneself and others only in virtue of a prior individuation of oneself and others as falling under the concept *person*.¹⁸⁶

This indicates, I think, that *person* is the fundamental concept by which individuals like ourselves are classified and that ‘I’ is an individuative term only in the sense that it picks out the person who is the user of the term.

If *person* is the primitive concept for the attribution of P-predicates, individual persons cannot be understood as pure mental entities. The reason is that, if the accessibility of a pure mental entity is possible only through the first-person perspective, then I cannot carry out the necessary identification of other subjects of experience for attributing P-predicates to myself.¹⁸⁷ Furthermore, nor can I

¹⁸⁵ See Chapter 7, section 7.5.
²⁸⁶ See chapter 5, section 5.5 and chapter 7, section 7.5. Firstly, a P-predicate’s univocal meaning in first-, second- and third-person ascription is intelligible only if the other-ascription of the P-predicate is due to a prior classification of the subject of the ascription as being a person, and that a certain state is a mental state only if there exists a person who is conscious of the state as a mental state.
carry out the necessary identification of other subjects of experience through identifying a subject of experience, other than myself, as the subject of experience who has the experience E, or who stands in the same causal relation to the body B as my experiences stand to body C. The reason for this is that we do not have any right to talk about experiences being mine in such a case, since the ability to identify other persons is a necessary condition for self-ascription of experiences.

As I have claimed already, this indicates that the only way that we can make sense of the ascription of P-predicates to oneself and others is through taking person to be the primitive ontological category, and ground the P-predicates on the basis of the observable behaviour of other persons. It is the observable behaviour of other persons that functions as the adequate criteria of application of a P-predicate in the sense that someone who is learning the P-predicate learns it as being applicable to objects categorised as persons with certain behavioural features typically associated with the P-predicate. This means that person is conceptually prior to ‘consciousness’.

Secondly, the fact that the adequate criteria for the application of a P-predicate are given through the observational behaviour of other persons makes it reasonable to suppose that persons must have material properties, contrary to Cartesianism and the Subjective View, since observable behaviour cannot be understood in any other way than observable physical, or bodily behaviour.

Furthermore, it is reasonable to suppose that a person, as the agent of actions, who tries to satisfy his intentions, wishes and dreams in the performance of his actions, must be a part of a system consisting of material objects related by time and place, since the carrying out of certain intentions, wishes and dreams is possible only if the person is interacting with the empirical world. The same is equally true of the person as the subject of experiences. To acquire beliefs, knowledge and sensations from the world, it is reasonable to suppose that a person must be a part of a system of material objects related by time and place, since it is the impinging of stimulus from the empirical world that provides the person with the information grounding his beliefs, knowledge and sensations. That is, it is reasonable to suppose that a person has to have certain properties in common with objects of the empirical world if he is to be able to interact with objects of the empirical world, since it seems reasonable to suppose that he can only interact with objects of the empirical world if he himself is a part of the spatio-temporal system of the empirical world.

---

188 See chapter 7, section 7.5.
These properties that persons must have in common with objects of the empirical world must be physical. The reason for this is that the applicability of the P-predicates to a person relies upon the existence of some special "empirical facts", so that the non-existence of these facts would imply the non-applicability of the P-predicates.\textsuperscript{189} For instance, if a person has the intention of reading a book that is in the college library, the person has to take the book from its shelf with his hands, sit down on a chair, and with his eyes follow the physically implemented letters of the pages of the book. In this sense, the true attribution of P-predicates of individual persons is due to the fact that persons are objects of which it makes sense to attribute M-predicates, such as 'is in the college library', 'weighs so-and-so' and 'has eyes' which implies that persons have certain physical characteristics as well as psychological characteristics.\textsuperscript{190} For instance, as David Wiggins has claimed,\textsuperscript{191} conceptually speaking a typical P-predicate such as 'remembering' is a particular kind of M-property, since having a veridical memory, i.e., the existence of the P-property memory in a certain person, requires, besides the purely inner representational content of an earlier event, the existence of a neuro-physiologically implemented causal relation between the person's apparent memory and the event which the apparent memory is a memory of.\textsuperscript{192} A person's perceptual experiences of external objects are also dependent upon the existence of M-properties, since it seems reasonable to claim that a person can only see an external object $o$ if $o$ is observable from the place where a certain head is located, and if certain functioning and open eyes are oriented towards $o$, so that $o$ falls in the person's visual field such that the person can receive visual information from $o$. If these empirical facts do not obtain, then the person would not perceive the object $o$, but have a misperception or hallucination of an object $o$.\textsuperscript{193} In this case too the attribution of a P-predicate to a person turns out to depend, not upon the existence of a pure P-property of the person, but upon an M-property of the person. Similar empirical facts, as the one obtaining in a visual experience, must also obtain if a person receives auditory, tactile, taste, smell and balance information. Furthermore, it is of fundamental importance that empirical facts also obtain when a person performs an action. For instance, the intentional kicking of a ball requires some

\textsuperscript{189} Strawson, \textit{Individuals}, p. 90.

\textsuperscript{190} \textit{Ibid.} p. 104-105

\textsuperscript{191} David Wiggins, 'The Person as Object of Science, as Subject of Experience, and as Locus of Value', pp. 64-66.

\textsuperscript{192} \textit{Ibid.} p 209.

\textsuperscript{193} \textit{Ibid.} p. 65.
of the former facts of sensory receptors, but also facts about how a certain material entity, presumably the leg of a person, hits the ball in a certain angle and at a certain speed with the effect that the ball changes its location.

8.8 Psychological Substantialism

So far this elucidation of the person as being an object of which both P-predicates and M-predicates are applicable does not consist of a difficulty for the Psychological Non-Criterianist. A Psychological Non-Criterianist does not need to deny that the material parts of a person play a significant role in the acquisition of beliefs, sensations, knowledge and the performance of actions. What it needs to deny is only that a person is essentially embodied. This point has been stressed by a recent Psychological Non-Criterianist account; ‘Psychological Substantialism’.

According to Psychological Substantialism it is the case that (i) persons are psychological substances that are non-essentially material without being essentially immaterial, and (ii) that person is a “basic sort” in that it refers to a substantial kind which is distinctively psychological in character. In this sense, Psychological Substantialism tries to maintain a view of persons in which an individual person is distinct from all physical properties but nevertheless can have physical properties in the same way as he has mental properties.

In the ordinary case persons are objects that are embodied. You and I are both persons with bodies which enables us to participate in the empirical world. In this sense, both you and I know that we are objects of which it is intelligible to attribute M-predicates. But, according to Psychological Substantialism, neither you nor I can be understood as being identical with our bodies. Firstly, it seems

---

194 It is easy to understand why not all of the facts about sensory receptors need to obtain if a person is to be able to kick a ball. If a person is blind, he cannot see the ball, but still kick it. And it seems plausible to claim that you do not necessarily need the sense of taste, the sense of smell or the auditory sense to be able to kick a ball.


197 Lowe, ‘Real Selves; Persons as a Substantial Kind’, pp. 104-107, and *Subjects of experience*, pp. 5-203.
to be plausible to assume that a typical person should possess some form of self-
knowledge in which he knows that he is the unique subject which is F. For in-
stance, if a person is having a painful headache it seems to be the case that the
person should know that he is the unique individual who is having the headache.
The reason for this is that persons have a certain kind of privileged access to
their own present conscious thoughts and experiences in that the ownership of
mental states cannot be in error. 198 I cannot be in error concerning whether I am
the owner of, for instance, a pain which I refer to as 'this pain', since the very
fact of my being able to refer to that particular pain is in virtue of knowing that
the pain is mine. In this sense, 'This pain is my pain' is something which I seem
to know a priori. 199 While it seems reasonable to suppose that a person
necessarily knows of any of his present conscious thoughts and experiences that
he is the unique subject of that thought and experience, the same does not seem
to be true of the bodily characteristics of the person. It seems, for instance, not to
be inconceivable that one single experience is dependent upon more than one
body. 200 One such example is a Strawsonian subject of experience, i.e., a subject
of visual experience, S, and three different relevant bodies: A, B and C, such that:

(1) Whether the eye lids of B and C are open or not is causally irrelevant to
whether S sees; but S sees only if the eye lids of A are open. And if an operation is
performed on the eyes of A, the result affects S's sight, but not if an operation is
performed on the eye lids of B and C. (2) Where A and B may be, however, is quite
irrelevant to where S sees from, i.e. to what his possible field of vision is. This is
determined only by where C is. So long as C is in the drawing-room and the cur-
tains are drawn, S can see only what is in the drawing-room (...) But (3) the direc-
tion in which the head and eyeballs of A and C are turned is quite irrelevant to
what S sees. Given the station of C, then which of all the views which are possible
from this position is the view seen by S, depends upon the direction in which the
head and eyeballs of B are turned, wherever B may find himself. 201

Furthermore, it does not seem to be inconceivable that two selves, or persons
share one and the same body either. Consider for instance cases such as split-
brain patients, a pregnant woman with a foetus, Siamese-twins and cases of
multiple personality like Dr Jekyll and Mr Hyde. 202 In such cases a person
knows that a certain body belongs to himself or herself, but it cannot be known

198 Lowe, Subjects of experience, pp. 189-191.
201 Strawson, Individuals, pp. 90-91.
by the subject whether that body uniquely belongs to himself or herself and not simultaneoulsy to another subject. Finally, it is imaginable that a person could be in a state of complete sensory deprivation such that the person does not have any knowledge at all concerning what body it possesses, even though it might be true that it must possess some body or other.

Given the correctness of these case, then, it seems plausible to conclude that we can know, *a priori*, that a person is not identical with his body. A person, though being associated with a particular body under normal circumstances, is a substance to which it is appropriate to attribute mental features, since it is "the true subject of psychological predicates".\(^{203}\) In this sense, a person is a psychological substance distinct from the biological or physical substance which is the person's body. Persons are, in this sense, psychological substances and the literal attribution of physical characteristics to a person are ascribable to the person in virtue of the fact that the "self's physical characteristics 'supervene' upon those of its body".\(^{204}\)

Another argument for the same conclusion is that the concept *person* entails that persons are distinct from any material object.\(^{205}\) Regardless of whether one understands 'material object' as *parcel of matter* or as *living organism* it seems to be reasonable to hold the view that a material object is essentially material in the sense that the object ceases to exist if it ceases to be material.\(^{206}\) In this sense, it is inconceivable to imagine a parcel of matter, such as a lump of gold, or a living organism, such as a man, not being constituted by material parts, even though one might, in the case of living organisms, imagine that an individual who once was a living organism becomes constituted by synthetic parts.\(^{207}\) But the same is not true of the concept *person*. It is not *a priori* the case that persons must be material objects. As, for instance, E. J. Lowe says, the concept *person* is the concept of an object who essentially

acts and perceives, and knows that it does so: it is a perceiving, self-conscious agent, or, alternatively, an active, self-conscious percipient.\(^{208}\)

Persons are considered to be the source of intelligent activity in the sense that if we are in doubt whether something is a person, what we are looking for is


\(^{204}\) Lowe, *Subjects of experience*, p. 37.

\(^{205}\) Lowe, *Kinds of Being*, pp. 109-137, and *Subjects of experience*, p. 34.


\(^{207}\) Ibid. pp 107-108.

\(^{208}\) Ibid. p 109.
some form of evidence of intelligent activity stemming from the object. Since the concept of a person is so different from the concept of living organism or parcel of matter, and that we know that there is a close connection between sortal concepts and criteria of identity for objects falling under the sortal concept, there are no reasons whatsoever to suppose that personal identity should share the same criterion of identity as material objects. That is, the concept person is a psychological concept singling out objects in virtue of the objects having psychological properties, not material properties.

Though persons are psychological entities one should not, according to Psychological Substantialism, think that persons cannot have physical characteristics. It is even the case that persons in the ordinary case know that they are objects that are embodied. You and I, for instance, know that we have a body which enables us to participate in the empirical world. In this sense, both you and I know that we are objects of which it is intelligible to attribute M-predicates. If we did not have that knowledge of ourselves neither one of us would be able to have an experience of a localised sensation or be able to move a physical part "at will". I must, for instance, be able to individuate my hand either as being the hand of which I am about to move at will, or as being the hand in which I have a certain sensation. Since volitions and sensations seems to be individuated through their phenomenal character which contain, at least partly, the physical part as a constituent it seems to be circular to think that I can individuate the physical part through firstly individuating my volition or the localised sensation. The idea that one must make an implicit reference to oneself, for instance 'my left foot', for being able to refer to the foot of oneself seems to be odd, since I simply do not seem to need to specify to myself whether some foot is mine to be able to move it at will.

Since persons can know that they have physical parts it seems reasonable to suppose that persons also can be attributed physical characteristics. This is also what Psychological Substantialism claims. For instance, 'I weigh so and so' attributed to myself should be understood as literally being a sentence about myself. It is a sentence which is true if and only if certain empirical facts are true of me, such that I am a spatial object which weighs whatever the statement

209 Lowe, Kinds of Being, p 110.

210 Ibid, p 112.


Psychological Non-Criterianism

says I weigh. In this sense, persons are not essentially immaterial objects. A person can be an extended object with spatial parts without being identical with their spatially extended parts.\(^{214}\) But the physical characteristics truly ascribable to the person is something the person possesses in virtue of having a body which has the physical properties. Describing a particular body as being mine is done in a derivative way. It is only through the intimate feeling of a certain body being directly effected by my “volitions” and where I can localise my sensations that I am able to determine that a certain body is my body. In this sense, a certain body can qualify as being mine only in virtue of being subject to what properly qualifies as mine; my mental states.\(^{215}\) In this sense, persons have physical properties in a derivative way only and should be understood as simple substances without substantial parts.

A consequence of the idea that person is a psychological concept and that persons should be understood as distinct from their bodies is that the term ‘person’, according to Psychological Substantialism, refers to a “basic” substantial kind which is psychological in character.\(^{216}\) Persons are, in this sense, concrete thinking substances\(^{217}\) or continuants who can persist through qualitative change\(^{218}\) and who comprise a natural psychological substantial kind which lacks informative criterion of identity.

By claiming that Person is a “basic sort” a Psychological Substantialist maintains that there simply cannot be found any criterion of identity for objects falling under the concept person. The reason for this is that the only two candidates for the criterion of identity for persons are inadequate. Firstly, physical continuity of any kind is inadequate on the ground that person refers to a sort whose instances are not essentially material. If it is the case that parcels of matter and biological organism do contain a constraint which entails that anything falling under the concepts is essentially material and that person does not consist of a necessary material condition for anything falling under the concept, then personal identity simply cannot consist in the continuity of physical parts.\(^{219}\) Secondly, even though psychological continuity does determine an important

\(^{214}\) Lowe, Subjects of experience, p. 36.

\(^{215}\) Ibid. pp. 36-37.

\(^{216}\) This account of Psychological Substantialism rests on the plausible idea that one place can at the same time be occupied by several objects as long as the objects are of different sorts.


\(^{218}\) Lowe, Subjects of experience, pp. 2-3.

\(^{219}\) Lowe, Kinds of Being, pp. 111-112.
condition of personhood according to Psychological Substantialism, it cannot be understood as the criterion of identity for persons. The reason for this is the point already presented in chapter 5 that it is impossible to individuate an individual mental state or a succession of mental states without individuating the state or the succession of states as being mental states of a particular person.\footnote{Lowe, \textit{Kinds of Being}, pp. 131-133, and \textit{Subjects of experience}, pp. 25-32. See also chapter 5, section 5.5.} Persons are, in this sense, prior to their mental states and the identity of persons is “primitive or ungrounded”.\footnote{Lowe, \textit{Subjects of experience}, p. 41.}

By claiming that \textit{Person} is a substantial kind a Psychological Substantialist maintains that ‘person’ is a term that has persistence-conditions associated with it which are distinct from the persistence-conditions of other kinds. Even though ‘person’ lacks an informative criterion of identity it is still the case that objects falling under the term persist in virtue of conditions which are specific for objects of the kind referred to by ‘person’. This is true even if it is the case that persons normally are embodied, since the person is distinct from the body. In the same way as ‘bronze-statue’ refers to a sort which is distinct from the sort referred to by ‘piece of bronze’ on the ground that a bronze-statue and a piece of bronze can occupy the same places at the same times without being identical,\footnote{That ‘piece of bronze’ and ‘bronze-statue’ refer to different sorts is obvious, since a piece of bronze might continue to exist even though the bronze-statue no longer exists.} so also ‘person’ can be understood as referring to a distinct kind of its own. A person might occupy the same places at the same times as a certain body without thereby being identical with the body. ‘Person’ refers to a psychological substantial kind of object in the world whose characteristic features are psychological or mental. In mentioning the existence, or describing the activities of persons one must use a psychological vocabulary in which higher cognitive capacities are attributed to the persons. And even though it is true that persons might, as in the case of ordinary human persons, be embodied in biological organisms, one should not think persons are their biological bodies. In this sense, persons are the genuine subjects of mental states. It is persons who think, love, hate, experience fear, experience lust, have emotions and so on. Furthermore, in describing the development and growth of persons, and in explaining actions and states of persons one must incorporate psychological, or socio-psychological laws. The reason for this is that persons just cannot be understood as subjects of their biological functions, but must be thought of as “products” of intellectual capacities involving socio-cultural interaction and organisation such as the development of
a language, technology, social structure, etc.\textsuperscript{223} Persons are, as E. J. Lowe analogously claims, similar to a lump of clay which can be formed in many different ways.\textsuperscript{224} In creating a vase, a potter takes a suitable material for creating the vase, i.e. some lump of clay, and creates a new substantial object distinct from the clay. The fact that the clump of clay is suitable for creating a vase is due to both the inherent properties of the clay and the properties of the creator of the vase, such as the creator having hands by which he can form the clay. In a similar way, persons form biological ‘clay’ into new persons, where a person is a new substantial object distinct from the suitable biological ‘clay’. The suitability of the biological material in the creation of persons is not only due to biological features of the biological object, but also “a function of the creative processes available to us given our own particular limitations”.\textsuperscript{225} The most important point here, though, is that one should not think that the biological material, though being suited for embodying a person, has a distinctive biological propensity for being a person, just as a lump of clay need not be thought of as having an internal distinctive propensity for being a vase.\textsuperscript{226} That is, persons are not biological organisms, but psychological substances who can be embodied in a biological substance.

\textbf{8.9 Criticism of Psychological Substantialism}

Though I am very sympathetic to Psychological Substantialism, especially the idea that persons are substances which we refer to with the basic sortal term ‘person’, I do not think it provides the correct understanding of what it is to be a person. The general scepticism I have regarding Psychological Substantialism is the insistence on \textit{person} being a psychological kind consisting of psychological substances. I will argue that there are strong reasons to suppose that persons are essentially embodied, even biologically embodied in a single body. I will first try to answer the alleged case of differentiation between persons and their biological bodies appealed to by the Psychological Substantialist. Secondly, I will criticise the idea that person is a psychological kind on the ground that the only

\begin{itemize}
  \item \textsuperscript{223} Lowe, \textit{Subjects of experience}, pp. 47-51.
  \item \textsuperscript{224} Ibid. pp. 50-51.
  \item \textsuperscript{225} Ibid. pp. 50-51.
  \item \textsuperscript{226} Ibid. p. 50.
\end{itemize}
available account of the fact that persons have higher cognitive states is one which understands persons as animals.

(i) Psychological Substantialism appeals to the possibility of Strawsonian subjects and the fact of split-brain patients and multiple personality disorder as showing that persons are distinct from their biological bodies. But, it is, on the one hand, questionable whether Psychological Substantialism is correct in maintaining that it is conceivable that there might be Strawsonian subjects, and, on the other hand, that split-brain cases and multiple personality disorder really point in favour of the view that more than one person can be associated with one body.

(a) Firstly, it does not seem to be a contingent fact that different experiences are dependent upon one and the same body. Consider, for instance, a person's belief- and action-grounding experiences. Suppose we have a Strawsonian subject S, whose experiences depend upon A's eyes, B's head and C's bodily location, who believes that the experienced object \( O \) is on the left side of him. The question is, now, what is it that S believes? Does S believe \( O \) to be to the left of A? To the left of B's head, or maybe to the left of C's point of location? As I understand it, all of these possibilities must be open for S, but none of them seems to be satisfactory. On the one hand, since perceptual information, if it is to be further processed, must impinge upon a retina it seems reasonable to assume that S must think \( O \) to be to the left of A. But at the same time, since the visual field of the eyes are determined by the direction of the head (for instance, what is behind me while I am typing this sentence is not part of my visual field) it seems that S should take \( O \) to be to the left of B. Lastly, since the visual field also depends upon the location of the body, (for instance, College Hall is not a part of my visual field while I am typing this sentence), it seems that S could understand \( O \) to be to the left of C. This means, then, that, if a Strawsonian subject occurred, we simply could not know, or determine, what that subject meant by the claim that an object \( O \) is to the left of him.

Furthermore, not only is it the case that we cannot know what S means by claiming that \( O \) is to the left of him, neither can S himself know what he means. There simply cannot be anything in the experiences of S which determines what S actually means by claiming that something being to the left of him. To claim,

\[\text{Thought Psychological Substantialism does not put the argument in terms of a person's belief- and action-grounding experiences, these are the terms my counter-argument will be formulated in. Using belief- and action-grounding experiences is not central to my argument. I could equally well have stated my argument only in terms of direct visual experiences. The advantage of using belief- and action-grounding experiences is that the unintelligibility of a position such as Psychological Substantialism becomes more apparent.}\]
as someone might want to do, that S believes O to be to the left of his *point of view*, is only to confuse things even more. If it is true that persons can be embodied, then there just cannot be something which could be called ‘the subject's point of view’ which does not depend upon the location of a certain body. What is true of ‘to the left’ is also true of other spatial concepts which S tries to apply. In no situation can S determine whether an object O is to the left of him, to the right, in front or behind. For S to be able to fix the spatial location of an object O, it is necessary that the information-processing systems causing the visual perception of O are co-ordinated, so that they give S only one possible visual field. This is what the having of one single body enables us to have.

Furthermore, since these beliefs of the subject are to structure his actions, the complete lack of determination of orientation of objects in the world would deem the Strawsonian subject to be nothing but a passive observer of events. Suppose someone, at the place of C, tries to hit C in the face. If S, who sees the blow coming, now wants to avoid this painful moment, he must turn C's head away from the blow and then turn B's head to look for a possible way to escape. But, since B is at a place where C is not, this seems to be of no help at all. What B can “see” is irrelevant for the escaping of C. But this means that it is of essential importance to an ordinary person's perceptions and actions, if the person is to be able to act correspondingly to his beliefs and perceptions, that they stand in a causal relation to a single body. Consequently, not only is it necessary that ordinary persons are embodied in a body, it is also necessary that an ordinary person at a particular time is embodied in one single body.

(b) Secondly, split-brain cases does not straightforwardly point in favour of a plurality of persons being embodied in one single body. It is important to notice that a “plurality view” of these cases involves an interpretations which one does not need to accept. As Thomas Nagel has pointed out, there are at least three different plausible positions concerning how to describe and understand split-brain cases.229 Interpreting singularity of mind as equivalent to singularity of person, as we ordinarily do, we can on the one hand interpret a split-brain patient as two persons, since there are “two minds, one which can talk and one which can-

---


229 Thomas Nagel, ‘Brain Bisection and the Unity of Consciousness’, from *Mortal Questions*, Canto edition, (New York: Cambridge University Press, 1991), pp. 147-164. Actually, Nagel discerns five different positions concerning how to interpret and describe split-brain cases. I have omitted two of them because these two seem to me, as well as it does to Nagel, completely implausible.
The second interpretation also claims that we have two persons in the split-brain case, but that the patient is only one person as long as the two hemispheres function in parallel. The third possibility, and the one I will defend, is that there is only one person involved in split-brain patients, but that the contents of that person's mind "derive from both hemispheres and are rather peculiar and dissociated".

The reason for not accepting the "two-person-interpretation" of split-brain patients is that it has several problems associated with it. Firstly, there is nothing which points in favour of there being two persons in one single biological body as long as the split-brain patient is situated in normal circumstances. For most of the time the patient acts in ways which indicate that only one person is involved. The patient walks in a normal way, retrieves things from his visual and auditory field and communicates in a consistent way with other persons in the world. It is, then, only in very special experimental situations that the actions of the patient become inconsistent. Even if the patient is situated in an experimental situation, it is not necessary that we get inconsistent responses from him. If the patient is allowed to receive symmetrical information, for instance touch objects with both hands and smell objects through both nostrils, nothing indicates any inconsistency in the behaviour of the patient.

Furthermore, even though the Corpus Callosum has been severed, there are still nerve-fibres connecting the two hemispheres. These nerve-fibres connect the hemispheres through the brainstem and the more primitive cerebellum. As long as these connections are intact there is a plausible explanation of the ability of the patient to receive a unitary world view with his sensory capacities which is consistent with the assumption that only one person is involved in a split-brain case. As Nagel himself says, to suggest "that we are not in a position to ascribe all those experiences to the same person, just because of some peculiarities about how the integration is achieved" seems to be very strange. A third reason for interpreting a split-brain patient as only one person, even though it is a person with a certain dissociated mind, is that other persons who know the patient think of him as a single person. This is true of both the experimenter, as well as closer friends of the patient.

---

231 Ibid. p 155.
232 Ibid. p 155.
233 Ibid. p. 159.
234 Ibid. p. 159.
But it is important to notice that there are also certain problem associated with understanding the patient as one person. One of the problems is that it is difficult to understand what it would be like to be a person with a dissociated mind, and the second problem is that split-brain persons do not satisfy the normal assumption that a person has a completely experientially connected domain.\textsuperscript{235} Contrary to Nagel, who thinks the whole problem of the split-brain patients shows that "the ordinary conception of a single, countable mind cannot be applied to them [the split-brain patients] at all, and that there is no number of such minds that they possess",\textsuperscript{236} I think the problems associated with treating split-brain patients as one person, i.e., one mind, are less problematic than treating them as two persons.

The problem that it is difficult to understand what it would be like to be one of these patients is not particularly damaging for the assumption that we are only dealing with one person in a split-brain case. We have trouble understanding what it would be like to be other persons too if the persons or the circumstances differ too much from what is normal for us. Too much difference between interpreter and the person who is to be understood from the inside makes it impossible to reach a thorough understanding of what it would be like to be that person.

The second problem is not that difficult to answer either. Why should we expect that an experientially connected domain is a necessary condition for being one person? Though it is normally true that a person is experiencing his experiential field as highly integrated, it is imaginable that certain individuals might have experiences, for instance a particular thought and a particular emotion at the same time without being able to consciously connect these two experiences. And the same hold for experiences over time. Being able to consciously connect one's different experiences in a connected domain is not a necessary feature for being a person, though it is a typical feature of persons. Typically, persons have one single stream of consciousness, but, contrary to the tradition where singularity of stream of consciousness is taken to be equivalent with singularity of person, I think we should accept that there is nothing, in and of itself, which prevents the possibility of one person having more than one stream of consciousness going on at the same time. So, all in all I do not think that split-brain cases point in favour of a plurality of persons in one single biological body. What it shows is that there could be a plurality of streams of consciousness' in one person.

\textsuperscript{235} Nagel, 'Brain bisection and the Unity of Consciousness', pp. 160-161.
\textsuperscript{236} Ibid. p. 155.
(c) The same is true of cases of multiple personality disorder. In such a case, the exhibition of internally different, and even inconsistent, psychological traits and abilities over a period of time of one single biological body, as in the case of Miss Beauchamp, is understood as entailing that one biological body inhabits several persons at different times. The general reason for the plurality view is, I think, that we would not have thought there to be any problems at all in having different persons if the characters or personalities of multiple personality disordered patients would have been associated with different biological bodies.\(^{237}\)

This does not imply that one should understand these personalities as different persons. What it shows us is that the individuation of a personality, i.e., a set of characteristic features, does not exclusively pick out individual persons. For the most part we can single out a person via the character or personality of the person. The reason for this is that most persons have character traits of a certain consistency and coherence throughout their lives, and changes of character or personality features are often of a gradual sort. If a set of features change rapidly, then many features often remain intact during the change of character of the person. But we should not think of the individuation of a person via the personality of the person as an absolutely successful way of individuating individual persons.\(^{238}\) The reason for this is that a personality might be instantiated in several different persons, in the sense that several persons can be shy, witty, ambitious or aggressive. That is, a character or a personality is something general. Characters, or personalities are more like sadness or happiness. Nobody would be inclined to think that a person dies just because he is going from happiness to sadness, and the same is true of personalities. If someone changes his personality, he just does not literally die. He continues to live, but with another personality. As Williams points out, if we say of someone that "He has the same character as his father"\(^{239}\) it does not follow that the father's character is going to be destroyed by the fact that the son's "character is going to be ruined by the Army".\(^{240}\) In this sense, characters, or personalities, simply are not particulars like watches, balls or persons.\(^{241}\) In this sense, personality, if it


\(^{241}\) *Ibid.* p. 18. Williams makes the point comparing the generality of character or personality with a watch. As Williams puts it, if to say of someone that "He has his father's watch", then we are committed to holding that "if the watch is going to be pawned tomorrow, his father's watch is going to be pawned".
should have any individuative force at all, has to be individuated through a more fundamental individuation in which it is related to a certain biological body. This means, then, that one person can have different personalities at different times in the same way that one and the same person can at one time be happy and at another time be sad, or have different social roles at different times. That is, we have a way of understanding patients diagnosed with multiple personality disorder which is intelligible and consistent with common sense. A patient diagnosed with multiple personality disorder should be understood as a single individual person with a fragmented personality which makes the person’s personality change discontinuously with the consequence that the person sometimes can remember what he at other times cannot remember.242

(ii) It is, I think, questionable whether person should be understood as a psychological kind. The general argument for this is that it is doubtful that the capacity for having higher cognitive capacities is a necessary condition for me to persist over time. If this view of mine is correct, then it seems doubtful that one could call person a psychological kind. As I showed earlier, in ‘vegetable cases’ and ‘foetus cases’ our intuition leads us towards a view which holds that one and the same individual can survive without having any capacity for higher cognitive capacities. For instance, if I, as a consequence of a car-accident, ended up as a non-cognitive creature at the hospital, the intuitive response is to hold the view that it really is I who am lying in the hospital, not somebody distinct from me. This is also true of the foetus case. We have strong intuitions implying that it is not incorrect to describe my development and growth as entailing that I am identical with a certain foetus of the past. What seems to be preserved in these two cases is a biological body, and not a capacity for psychological functions of the individual, and this seems to indicate that, if person is a substantial kind as Psychological Substantialism claims, then person is primarily a biological substantial kind.

Furthermore, according to Psychological Substantialism, one should not think that a person’s ability to have higher cognitive capacities is due to some biological features of the object, even though biological features are, as in the case of human persons, suitable for these higher cognitive capacities. A person is thought to be distinct from the biological object and belongs to a psychological kind. But I think this is doubtful.

242 Williams, ‘Personal identity and individuation’, p. 18.
As Snowdon has pointed out it is implausible to suppose that a human animal
cannot be the subject of psychological attributes.243 This is so, even if one thinks
that persons are the primary subjects of psychological attributes. As Snowdon
points out, it surely is the case that some animals have evolved with the capacity
for having thoughts about themselves and the ability to refer to themselves with
the help of the first-person pronoun. To assume otherwise would be to imply
that the statement ‘I am an animal’ could not be a statement which could be set-
tled by empirical facts, and which would not express an empirical truth. But, as
a matter of fact, ‘I am an animal’ seems to express an empirical truth and there­
fore one would have to assume that animals can be the subjects of psychological
attributes and have psychological capacities. The question is how one should
give an account of how animals can have higher cognitive capacities? The only
possible account I can think of is one which attributes this capacity of animals to
the real essence of the animals, i.e., to the structure of the biological tissue of the
animals where a primary significance probably should be attributed to the cen­
tral nervous system of the animals. That is, an animal who is able to have
thoughts of itself and who can refer to itself has these capacities in virtue of
having a certain biological structure. But it seems unreasonable to suppose that
human beings’, (who definitively are animals with these higher cognitive
capacities), have the ability for higher cognitive states due to some form of
biological propensity, while at the same time think that persons’, at least human
persons’, ability to have the same kind of higher cognitive states is not due to
some form of biological propensity. Why think the same states have different
origins? What is reasonable, I think, is to hold that persons’ ability for having
higher cognitive states is founded in the same deep level structure as animals’
ability for having higher cognitive states. In this sense, persons’ ability for
cognitive and intellectual activity is a function of their having a certain
organisation of their biological tissue. Persons are, in this sense, not
psychological simple substances, but complex biological substances. Hence, one
should not understand person as a psychological kind, but as a biological
substantial kind.

243 P. F. Snowdon, ‘Persons, Animals, and Ourselves’, in The Person and the Human mind -
issues in ancient and modern philosophy, edited by Christopher Gill, (Oxford: Oxford
8.10 One interpretation and one remark

I have claimed, following Wiggins, that persons are physical beings on the ground that P-predicates are matter-involving. In this sense, every P-predicate entails the existence of some M-predicate. Let me clarify this account a little, since I think there is a certain unwanted tension in Wiggins elucidation of the relation between P-predicates (P-properties) and M-predicates (M-properties).

On the one hand, it is possible to interpret the matter-involvedness of P-predicates as the weak thesis that the attribution of a P-predicate to a person implies not only the existence of a P-property of the person, but also an M-property of the person. A second possible interpretation of the same claim is the strong claim that P-predicates (P-properties) are a kind of M-predicates (M-properties). For instance, when Wiggins says that “a purely conceptual (though not necessarily purely a priori) inquiry discloses to us that the P-property of remembering is necessarily also a rather particular M-property” it seems that he wants to uphold the idea that, properly speaking, there are only one kind of predicates, the M-predicates.

But this strong interpretation of the relation between P-predicates and M-predicates is something we should try to avoid if we want to keep our account of what is essential for personhood in accord with how persons are, ordinarily speaking, presented to us. This is most obvious in cases of first-person reports of the existence of P-properties, such as a person's reports about his own experiences, memories and intentions. When having an intention, or an experience from the first-person perspective, what is presented is the existence of a psychological state, not a material state. For instance, if I close my eyes and wish myself to have a beer, I do not experience some neuro-physiological state which realises my wish, but a wish which presents itself to me as a wishing for a beer. Supposing it is possible in the future to identify a certain mental state with a certain neuro-physiological state, this would not influence how the P-properties are presented to a person from the first-person mode of presentation. We still would experience, from the first-person perspective, P-properties and not M-properties. In this sense, P-predicates are not a kind of M-predicates, since they are an irreducible device for predicating the existence of certain states in which a person can find himself.

Thus, we should interpret the matter-involvedness of P-predicate as entailing the weak thesis that the true application of P-predicates implies, or entails that certain M-predicates are also satisfied. As Wiggins writes

What we are concerned with here is not the reducibility of psychological concepts to concepts that pull their weight in the sciences of matter, but rather with the fact
that the elucidation of psychological concepts seems to be essentially matter-involved.\(^{244}\)

It is in this sense that a philosopher who takes the pure first-person approach to personal identity is wrong when he assumes that it is possible to have P-properties without having M-properties, since the having of certain P-properties depend upon the existence of certain empirical facts, or M-properties.

But, of course, it might still be claimed that I have not shown that it is essential for a person to have material parts. Since my point of departure was that persons are agents of actions and subjects of experience, what I have established is only that it is essential for persons like us to have physical parts. I have not established that it is metaphysically impossible that a person can exist in a dis-embodied state. This is correct. No conclusive argument showing that persons are essentially physical has been presented. But, I think that what I have said points in favour of not understanding persons as being primarily psychological beings.

Let me end this discussion with a last point in favour of treating persons as being physically embodied in biological organisms. What I am thinking of is the idea that the meaning, or the content associated with a term is externally determined. This, I said, meant two things; firstly, the content of the thought one has when one is thinking about an object is determined not by purely qualitative description alone, but relates the content to the actual world, and secondly, the semantic bonding between a term and its object is not established by a pure qualitative description, but by the external surroundings one was in when acquiring the term. Now, if this externalistic account of meaning is true of the term ‘person’, then the meaning of the term ‘person’ is determined by an ostensive definition in which it is declared that, whatever is to be counted as a person, must have the same kind of deep level structure as the persons of the actual world. Now, since the only possible candidates by which the term ‘person’ could have been ostensively defined are human animals, this implies that persons are animals. That is, since ‘person’ refers to objects which necessarily have biological properties in the actual world, then even if we found out that certain individuals were immaterial and could behave as we do, they would not be persons.

---

\(^{244}\)Wiggins, ‘The Person as Object of Science, as Subject of Experience, and as Locus of Value’, p. 66.
9. THE PRIMITIVENESS OF PERSON
AND PERSONAL IDENTITY

9.1 Introduction

In this final chapter I will put forward an account of person and personal identity according to which person is understood as a basic sortal concept picking out a substantial sort of object in the world. Individual persons are, according to this account, objects of certain biological sorts who endure over time. This account is Non-Criterianist in claiming that personal identity is primitive in relation to the underlying empirical continuities accompanying it, but it is a Non-Criterianist position which avoids positing persons as being primarily psychological beings. It is, then, an account which maintains that personal identity is a primitive further fact but where persons are understood as biological objects in an empirical world. The defence of this view consists in establishing that five consequences of the claim that 'person' is a basic sortal term are consistent with earlier established conclusions of the work and that the sortal concept person has an irreducible biological character. In doing this I will present and defend my position against certain counter-arguments. I conclude that the best available account of person and personal identity is one which (i) focuses upon persons as comprising a sort of enduring object which is biological in character, and (ii) that the conditions for being a person and remaining identical over time as a person are primitive in relation to the empirical facts that underlie them.

9.2 The basic sortal concept person

The apparent fact that it is impossible to specify non-circular and informative conditions for personal identity, or identity for objects like us, makes it plausible to maintain that person is a basic sortal concept. By the claim that Person is a basic sortal concept I mean that it is true of the sortal concept person that:

(i) person is a primitive concept by which objects are individuated,

(ii) personal identity is primitive in relation to the empirical conditions that underlie it,

(iii) persons are objects that endure over time in the sense that individual persons persist in virtue of being wholly present at each time of their existence,

(iv) 'person' is a substance sortal term in that it picks out an essential feature of individual persons,

(v) 'person' refers to a substantial sort Person of objects in the world.

(i) The primitiveness of the sortal concept person: That person is a primitive concept is shown by the fact that it is necessary that individuals like us, that is individuals like you and I, individuate themselves as falling under the sortal concept person. This does not mean that we are not essentially human beings, but that we must individuate ourselves as being persons before we can individuate ourselves as human beings. The only possible way in which it can be established that human being is an essential property of objects like us is through the use of first-person pronouns and that first-person pronouns require, for their intelligibility, a prior understanding and application of the concept person to oneself and others. Only if I can individuate other objects as persons, as objects satisfying M-predicates and P-predicates, can I think of and experience myself as an individual with consciousness and self-consciousness. In this sense, a mutual and reciprocal preparedness for attribution of personhood to oneself and others is necessary for our concept person to have the significance it actually has in our conceptual framework.188

A further reason for treating person as a primitive concept is that the concept of a person signifies a unified individual consisting of both mental and physical states and that neither the mental nor the physical states of a person can be

188 Ideas similar to this have been endorsed by Hidé Ishiguro, 'The primitiveness of the Concept of a Person', in Philosophical Subjects: Essays presented to P. F. Strawson, ed. Zak van Straaten, (Oxford: Clarendon Press, 1980).
primitive, since neither can be individuated without using, or presupposing the concept *person*. That mental states cannot be primitive I have already shown when I claimed that an individual mental state is always related to a subject who is the bearer of the mental state. But, it is equally true that physical states of persons need to be related to the person whose physical states they are if these physical states are to be individuated.

The reason is that in identifying the body, and not the person, as the primitive unit, a theorist must claim that it is possible to determine that a certain body \( B \) is the body in which experiences occur, or alternatively, that the body \( B \) is causally responsible for the occurrence of experiences, and then conclude that the experiences are the experiences of a certain person \( P \). It is important that the experiences should belong to a particular person \( P \), since otherwise the body \( B \) would be the body of which all experiences are dependent. But this is not something one would like to be the case, since it is a commonly accepted fact that there are some experiences which are mine and some experiences which are yours and so on.

But the only possible way in which it is determinable that a certain body \( B \) is the body of a person \( P \) without presupposing the concept *person* is through an individuation of a certain class of experiences as being causally dependent upon, (or even identical with), a certain body, (or body-state), \( B \). But now, even if we accept, as I do not, that it is possible to have the distinction between experiences belonging to different persons without presupposing the concept *person*, the possibility of determining that a certain class of experiences \( e \) depends upon the body \( B \) relies upon the use of the concept *person*. The reason is that it is not possible to find a way of singling out the particular class of experiences \( e \) without presupposing that this class of experiences is, as Strawson puts it, ",my experiences" or 'the experiences of some person'".\(^{189}\) The reason for holding that it is impossible to determine that a class of experiences depend upon a certain body \( B \) is that experiences, even the experiences of bodily states of a person, are individuated in relation to the person whose experiences they are. As I, following E. J. Lowe, already have claimed, determining the presence of a mental state requires a prior determination that there exists an object which can experience the mental state.\(^{190}\) It is only with the help of terms indicating the presence of a person, such as ‘I’, ‘my’, ‘you’, ‘she’, ‘person’, etc. that we can individuate the class of experiences which, so to speak, are to be said to be determined by a cer-

\(^{189}\) Strawson, *Individuals*, p. 97.

\(^{190}\) See chapter 5, section 5.5.
tain body. In this sense, a body is determined as being my body, or the body of some person. That is, it is the case that both mental states, or consciousness, and bodily states cannot be individuated except as the states of some previously individuated person P. Consequently, the concept person can neither be analysed as the subject of experiences e caused by the body B, nor the subject of experiences e, but must be understood as a basic primitive concept which picks out unified objects who possess both mental properties and physical properties.

(ii) The primitiveness of personal identity: That 'personal identity' also should be understood as something primitive in our thoughts about the world is grounded in the fact that no Criterianist account has succeeded in specifying non-circular and informative necessary and sufficient conditions for personal identity. This claim, then, is a plausibility argument. Since the best available accounts of what would constitute the criterion for personal identity are inadequate, it is reasonable to infer that personal identity is a primitive concept that cannot be specified in any more basic terms. Though it might be the case that personal identity is associated with some form of empirical conditions in the sense that, if personal identity obtains, then these empirical conditions also obtain or vice versa, it is still the case that the obtaining of these empirical conditions conceptually rely upon a prior employment of 'personal identity'. Thus, even though it might be possible to determine conditions for personal identity, it is not possible to specify a non-circular and informative criterion for personal identity.

(iii) Persons as enduring objects: That personal identity is a basic fact is also supported by the fact that persons must be understood as enduring, rather than perduring, objects. As I have claimed, understanding persons as primary substances, or continuants, is ontologically prior to understanding them as four-dimensional objects. We simply cannot describe objects and events of the four-dimensional ontology without defining the terms of the four-dimensional language in relation to an already presupposed three-dimensional language referring to three-dimensional objects. This three-dimensional view is a part of the concept person, i.e. the concept person is only applicable to enduring objects. Personal identity between a person x existing at t and a person y existing at t* is strict numerical identity between x and y. If we combine this fact with the fact of the sortal dependency of individuation it becomes obvious that personal identity is, at the primitive level of description of our persistence over time. In individuating an object as being a person at a moment we also imply that that

---

191 See chapter 3, section 3.9.
person is strictly identical with some person existing at another time. But
claiming this is not different from claiming that personal identity obtains
between an individuated person and a past or future person and, hence, we
already employ an understanding of ‘personal identity’ in individuating persons.

(iv) ‘Person’ as a substance sortal term: I have already claimed that we have
reasons to assume that ‘person’ is a substance sortal term. The way we use
the term ‘person’, both when it comes to first-person usage and third-person
usage, entails that being a person is a property an object cannot cease to possess
without thereby ceasing to exist. ‘Person’ is, in this sense, similar to natural kind
terms such as ‘horse’. Both ‘horse’ and ‘person’ apply to the longest prolonga-
tion of objects falling under the term.

(v) Person as a substantial sort: Related to the idea that ‘person’ is a substance
sortal term is the idea that Person is a substantial sort of object. By a sort S
being a substantial sort I mean that objects falling under the sort S have a certain
mind-independent unity or structure which explains that they fall under the sort
in question. More formally, S is a substantial sort iff

(i) \( S \) is a substance sort (i.e., something being of the sort \( S \) cannot cease
to be of sort \( S \) without ceasing to exist),

(ii) \( S \) is a sort of enduring objects,

(iii) there exists an essence, or a set of properties, \( E \) such that, necessar-
illy, something is of sort \( S \) if and only if it has the essence \( E \).

I will argue that Person is a substantial sort according to this characterisation
due to the facts that (i) ‘person’ refers to a substance sort, and (ii) persons are
enduring objects that have a certain real essence. We already know that the first
condition is satisfied, since ‘person’ is a substance sortal term. The second fact
will be argued more thoroughly in section 9.3-9.7.

The general idea, though, is that Person is a sort that is similar to the natural kinds Tiger and Horse. These sorts have a strictly natural, or objective character. Individual persons are not social constructions in the sense that we can conventionally determine when and where an individual person begins and cease to exist. There is a real and natural unity in the individual persons themselves, a
unity which is prior to any conceptual structure formed by human mind. For
instance, an individual horse \( a \) belongs to the sort Horse in virtue of having cer-
tain properties. These properties, whatever they are, are what explains that the
horse \( a \) belongs to the sort Horse and not Cow. Assuming that \( a \) looses the
properties in virtue of which he belongs to Horse is the same as assuming the

\[192\] See Chapter 7, section 7.5.
ceasing of existence of *a*, since it seems reasonable to think that a particular horse *a* ceases to exist when it ceases to be a horse. In this sense, 'horse' is applicable to *a* as long as *a* exists and when 'horse' no longer is applicable to *a*, *a* no longer exists. Hence, an individual horse is a substantial object which ceases to exist when it no longer belongs to the substantial sort *Horse*.

The same is true of 'person'. Since 'person' refers to a substance sort, ceasing to be a person is the same as ceasing to exist as an object of the world. That is, an individual person *b* belongs to the sort *Person* in virtue of having certain properties which are essential for his persistence as an object. If *b* ceases to have these properties by which he is a person, then *b* ceases to exist. Thus, *Person* is a substantial sort of enduring objects.

### 9.3 Person and natural kinds

The assumption is, then, that 'person' is a substance sortal term which refers to a substantial sort of enduring objects with a certain distinctive unity structuring their existence. In this 'person' is similar to biological natural kind terms such a 'tiger' and 'horse'. They all aim at picking out distinct sorts of objects in the world where a sort is understood as being comprised of objects with certain essential properties. Let me now try to establish that there really exists an essence which is necessary for something to be a person.

Firstly, the above mentioned similarities between 'person' and biological natural kind terms are not the only similarities between 'person' and biological natural kind terms. I have indicated at several places that the only plausible account of the concept *person* is to understand it as a concept derived from actual persons in the world. Our concept *person*, whatever else it should be applicable to, ought to be applicable to individuals like ourselves, i.e., objects of the species *Homo Sapiens* who conceive of themselves as being persons. This point can now be strengthened somewhat, since we have reasons to suppose that *Person* is a basic sort. We can apply the concept *person* to objects of our sort, since this is actually done. This application of the concept *person* is, I have claimed, primitive in the sense that an individual object who knows he is falling under the concept must presuppose the concept *person*. But, this seems to indicate that the only possible way of acquiring the concept *person* is through a direct relation to one or several particular instances falling under the concept. In this sense, the concept *person* is like a natural kind concept in that we have the concept in
question in virtue of standing in a particular causal relation to one or several actual objects falling under it.

Another feature that the term ‘person’ shares with biological natural kind terms is that it is associated with certain empirical law-like principles constraining the specific development and history of individual members falling under the term. For instance, an individual horse develops from being a foal to a fully grown up horse due to the existence of certain inherent biological processes in the object, and the same biological processes prevent the object from passing through certain changes. An important point with these law-like principles constraining the development and history of a horse is that they contain both conceptual and empirical elements. For instance, while it is a priori true that horses are biological organisms and hence must persist as a biological organism, it is an empirical matter which biological processes are required for a biological organism to be a horse and what biological processes are required for the preservation of a horse. Thus, the biological processes associated with the substantial sort Horse need not be conceptually transparent in an analysis of the term ‘horse’. To specify what it is to be a horse requires certain empirical investigations which, once carried out, constrain what objects in the world can belong to the sort Horse.

The same is true for persons. Individual persons have a self-sustaining unity due to the fact that persons are subject to certain law-like principles which inhere in the persons themselves. There are certain changes which are compatible with the preservation of an individual person, and some that are not. This line of thought is realist in the sense that persons are understood as having a certain distinctive structure prior to our classificatory involvement. It is not because we have the concept person that we are able to have experiences of

---

193 Now, someone might claim that this bundling together of the term ‘person’ with natural kind terms involves us in a contradiction, since it might be claimed that the most reasonable way of understanding the ‘law-like principles’ that are supposed to be associated with natural kinds is that they constitute, even though they are a posteriori, the persistence conditions for an individual member of the kind, where ‘persistence conditions’ should be interpreted as meaning the same as ‘necessary and sufficient conditions’. That is, according to this view, natural kind terms are associated with the possibility of specifying empirical necessary and sufficient conditions for being a member of the kind. But, since I have claimed that it is a distinctive feature of the term ‘person’ that it does not have a criterion of identity, it seems that I just cannot treat ‘person’ on a par with natural kind terms. Now, this argument overlooks the fact that it is possible to interpret ‘law-like principles’ in a less definitional way, and only require of the law-like principles that it determines a posteriori necessary and sufficient conditions of individual members of the sort or kind in relation to the individual members already being individuated via a more primitive or basic sort. And, that it is possible to specify circular a posteriori necessary and sufficient conditions for being a person is not something which I have any reasons at all to deny.
individual persons, but we have the concept person because of the existence of individual persons. In the same way that the existence of individual tigers and horses makes it possible for us to conceptually classify something as a tiger or a horse, we are able to have the basic sortal person because there are individual persons in the world and that these individual persons have a distinctive way of persisting.

But though 'person' resembles a biological natural kind term in that there are law-like principles determining possible alterations in individuals falling under the term, this does not imply that 'person' is a natural kind term. The reason for this is that 'person' is not the name of a species in the way that 'horse' is a name of a species. In this sense, though Person is a substantial sort, it is not a substantial kind like natural kinds such as Tiger and Horse. The difference between a substantial sort and a substantial kind is that the objects belonging to the former are subject to laws distinctive for a substantial kind, and not the substantial sort in question. That is, objects of a substantial sort have their existence and persistence-conditions in virtue of belonging to a substantial kind, but where the laws determining the possible lives of objects of the substantial kind entail a set of properties, a real essence, which explains that objects of the substantial kind belong to a substantial sort. In this sense, a person exists and persists in virtue of belonging to a substantial kind, but where the laws determining the possible lives of objects of the substantial kind entail a set of properties, a real essence, which explains that objects of the substantial kind belong to a substantial sort. But a person has also a mind-independent unity which explains the fact that he belongs to the substantial sort Person. In this sense, our concept person is construed in such a way that individuals from several different kinds could fall under person. In that way person is a cross-classificatory concept akin to a biological natural kind concept in that it picks out a mind-independent sort of object without picking out a determinate biological species.194 Person, then, is not a natural kind concept, but a cross-classificatory concept picking out a sort which consists of objects from different kinds.195

194 Wiggins, Sameness and Substance, p. 172.

195 In this way person is both similar and dissimilar to the concept vegetable. The similarity between person and vegetable is that both concepts a cross-classificatory concepts clustering together objects from different natural kinds. The dissimilarity of person and vegetable is, on the other hand, that vegetable is a concept that is clearly functionally defined in relation to human beings, whereas person should be understood as picking out a primitive substantial sort of objects.
9.4 Person as a biological sort

Though *person* is a basic sortal concept that lacks an associated informative and non-circular criterion, I still think it possible to specify some important principles associated with it.

(i) Individuals like you and I are persons.

This principle is true on the ground that it is individuals like you and I who are the paradigmatic instances of the concept *person*. *Person* is a concept which is formed around individuals like you and I.

(ii) If $x$ is a person, then $x$ is an animal.

This principle is true on the ground that we, the individuals whom 'person' is centred around, are animals. If something is a person, then, by necessity, it must be like *us*, human beings, in all relevant ways. This means that whatever 'person' refers to, it must refer to individuals who are similar to you and I. But, since it is a reasonable hypothesis that individuals like you and I are persons in virtue of our possessing a certain internal constitution causing us to have features normally associated with personhood, it is also reasonable to assume that the same constitution should be found in all persons. Since we are objects of a biological character, *animals*, this is a necessary condition for being a person.

(iii) If $x$ is a person and $x$ is an animal of the kind $K$, then all individuals of the kind $K$ are persons.

This principle is true on the ground that 'person' is a relational concept similar to natural kind concepts. A concept can either be an individual concept or a relational concept. If a concept is an individual concept, then an object falls under it wholly in virtue of the object itself having the properties associated with the concept. If a concept is a relational concept, then an object falls under it in virtue of the object standing in a certain relation to some other individual object which falls under the concept and which has the properties normally associated with the concept. Typical examples of relational concepts are concepts picking out natural biological kinds. An individual object belongs to a kind partly because it has certain internal features of its own, but also partly because it stands in a certain causal and evolutionary relation to other members of the kind in question. For instance, something is a horse because it has a certain morphological structure grounded in a deep level structure which is causally and evolutionary related to specific individuals that exemplify what is meant by the concept *horse*. It is the fact that a particular individual object $a$ has the same internal constitution as the typical instance of the sort that makes it appropriate to clas-
sify \( a \) as belonging to the same kind as the stereotype. But, it is important to point out that a certain degree of individual variation of the morphological as well as the deep level structure is possible. For instance, a horse typically has four legs. But, a three-legged horse is no less a horse than a four-legged horse. A three-legged horse is related to ordinary horses causally and through evolution, and it is, despite the lack of one leg, this fact of causal and evolutionary history of the three-legged individual horse which makes it a horse. The same is also true of certain variations of the internal biological code of individual biological animals. Some mutations of the deep level structure of biological organisms do not imply that the mutated individual ceases to belong to its original natural kind. What it implies is that the mutated individuals are natural variations of the natural kind. But they continue to belong to the natural kind in virtue of a causal and evolutionary chain leading back to a typical individual member of the kind. In this sense, something is a horse, not in virtue of having certain necessary individual features, but in virtue of standing in a causal and evolutionary relation to an individual who belongs to the kind horse.

The same is also true of persons. Certain variations of the important internal structure are allowed as long as the object is related to a typical instance of the concept \textit{person}. In this sense, a person need not satisfy all properties a typical person exemplifies. For instance, while it is reasonable that our notion of 'the typical person' involves the idea of an object with perception, memory, intentions, emotions, reflections and the capacity for social interaction with other objects with similar capacities, it does not seem to be required that all individual persons must have all of these properties. An individual person is a person in virtue of being a member of a kind whose real essence is such that it explains the possession of typical properties of persons. This means that the concept \textit{person} is a relational concept. And given that an individual member of that kind is a person it will follow that other members of the kind are also persons, even though they do not have all the features normally associated with being a person.

(iv) If \( K \) is a kind of animal whose members are persons, then the real essence of \( K \) would, if completely specified, give an account of the possession of a fully developed individual of the kind to have cognitive and social capacities similar to us, human beings.

This principle establishes, once again, that the concept \textit{person} is formed around us, individuals like you and I. One of the central features of individuals like you and I is that we have a mental and social life. It is in virtue of the fact that individuals like you and I have these higher mental and social capacities that the concept \textit{person} can be said to pick out a distinctive sort of object.
Though both you and I are persons with higher cognitive and social capacities, it is less than clear what capacities we actually must have in order to be persons. But, if one starts investigating what are the cognitive and social capacities that persons such as you and I possess, what one will end up with is some form of open-ended list of characteristics similar to the "capacity-clause" of the animal attribute view. That is, the abilities to perceive, feel, remember, imagine, desire, make projects, move themselves at will, speak, carry out projects, acquire a character as they age, are happy or miserable, are susceptible to concern for members of their own or like species... [note carefully these and subsequent dots], conceive of themselves as perceiving, feeling, remembering, imagining, desiring, making projects, speaking..., have, and conceive of themselves as having, a past accessible in experience-memory and a future accessible in intention..., etc.\textsuperscript{196}

Since we do not know what cognitive and social characteristics individuals like you and I possess, we cannot specify the necessary and sufficient conditions for being a person. What we can do is to create a list of important features normally associated with human beings, but where no individual capacity of the list is absolutely necessary, nor sufficient, for personhood. In this sense, the "capacity-specification-clause" make person into an open-ended concept.

Furthermore, it is plausible to assume that it is in virtue of being members of the species \textit{Homo Sapiens} that individuals like you and I have the capacities normally associated with personhood. For instance, to reflect, perceive, imagine, remember, etc., is something that I can do in virtue of the species \textit{Homo Sapiens} having a real essence which enables individual members of the species to have certain mental features.

I think it is a plausible empirical hypothesis that having a neuro-physiological structure, in some way or other implemented in a central nervous system, is a necessary condition for having a mental life, even if it cannot be understood as a sufficient condition for the exact content of the mental life.\textsuperscript{197} My mental sphere is, in this sense, partly a causally emergent feature of my neurophysiological structure. Furthermore, it is equally plausible to assume that the general structure of my central nervous system is due to the fact that I am a member of the kind \textit{Homo Sapiens}. I have a certain neurophysiological structure in virtue of being a human being that is causally and evolutionarily related to other individual

\textsuperscript{196} Wiggins, \textit{Sameness and Substance}, p. 171.

\textsuperscript{197} The reason that having a certain central nervous system does not seem to be sufficient for having a certain mental life is that it is plausible to assume that mental content is externally determined. The content of a thought is not exclusively determined by the neuro-physiological structure as such, but is dependent on the linguistic community, environment and social interactions in which an individual participates.
members of the same species. In this sense, it is reasonable to assume, as Animalism does, that the real essence of the person that I am is that of being an animal of the species Homo Sapiens. The same would be true if there existed persons other than human beings. If, for instance, dolphins are persons, then an individual dolphin is a person in virtue of belonging to a kind whose real essence is such that its members develop a neurological structure which causes the occurrence of mental capacities of individual members of the kind. In this sense it is in virtue of having the real essence of the natural kind that an individual animal is a person.

(v) If K is a kind of animal with a real essence which would, if completely specified, give an account of the possession of a fully developed individual of the kind to have cognitive and social capacities similar to us, i.e., human beings, then we have good reasons to suppose that animals of the kind K are persons.

This principle is, once again, relating the concept person to us. But this principle does not state a necessary condition for personhood, only a default principle of our concept person by which it would be possible to determine whether animals, other than ourselves, are persons. It is possible to interpret the required similarity to human beings of person in a liberal way and claim that it is an open question whether there are any kinds of animals, except human beings, that are persons. But if these animals are persons, it is in virtue of having a real essence similar to human beings in that it sustains higher cognitive and social capacities similar to those of human beings of a fully developed individual of the animal kind. For instance, it might turn out to be the case that dolphins and certain higher primates are persons. If these species of animals are such that a fully developed individual of the kind can perceive, feel, remember, have concern for members of their own or like species, can reflect upon their perceptions, feelings, memories, and experience their own existence through time as containing a past and a future, etc., then these animals most likely are persons. A more far fetched example would be a group of animals, absolutely distinct from human beings, from another planet satisfying the required characteristics for being a person. As long as the object is an animal, a biological organism, of a kind whose typical members possess the same kind of higher cognitive social capacities as human beings possess, then there are good reasons to suppose that that object is a person. But it is important that this principle is defeasible. The having of similar cognitive and social capacities of an animal of a kind K, unless it is a human being, only provides us with prima facie good reasons for assuming that animals of the kind K are persons. This conclusion can, though, always be
The Prim itiveness of Pers on and Personal Identity

199

defeated by further evidence that points against animals of kind K being persons. The reason for this is that person is an open-ended concept with certain constraints associated with it. As already stated, the cognitive and social capacities that persons normally are assumed to possess cannot be given a determinate specification, since we simply do not know what are the exact features that individuals like you and I possess. But since we cannot specify the exact mental and social conditions for personhood, we cannot specify a generally valid criterion of application for persons.

The last principle of my account of what it is to be a person is

(vi) $x$ is the same person as $y$ iff $x$ is a person and $y$ is a person and $x$ is the same animal as $y$.

This principle entails that sameness of person entails sameness of animal and that sameness of person-animal entails sameness of person. That is, in no circumstances is it possible that a person, who is an animal by necessity, might continue to exist without the animal also continuing existing. Further, in no circumstances is it possible that an animal that is a person continues to exist without the person also persisting.

The conditions that I have stated above make my view quite similar to the animal attribute view of personhood due to David Wiggins. In accordance with Wiggins' characterisation, I claim that the concept person is a cross-classificatory concept ranging over a certain class of biological natural kinds, i.e., animals of different sorts, where the sort enables individuals to have an unspecifiable number of higher cognitive and social capacities such as the ability to perceive, remember, imagine, care for individual of their own or like species, and are able to conceive of themselves as existing over time. But contrary to the animal attribute view of Wiggins, I explicitly claim that (i) it is not possible to specify necessary and sufficient condition for personhood, (ii) 'person' refers to the basic substantial sort Person which has a distinctive real essence, (iii) person is a biological cross-classificatory concept, and (iv) the concept person, though picking out a substantial sort, is attributable to certain sorts of animals in virtue of the real essence of the animal kind.

It is important to notice that my claim that person is a cross-classificatory concept ranging over different biological kinds entails two different claims:

(i) The epistemological claim: As far as we know, it might be the case that the species Homo Sapiens is not the only existing biological kind whose members

198 See chapter 7, section 7.2, and Wiggins, Sameness and Substance, pp. 171-173.
qualify as being persons. For instance, it might turn out to be the case that dolphins or some primates are also persons.

(ii) The ontological claim: In the same way as there could have existed mammals even though no actually existing kind of mammals existed, it might have been the case that there had existed other biological species than those actually existing that would have qualified as being persons. Given that person is determined in relation to the actual species of human beings (Homo Sapiens), a possible biological species would, if it had existed, been a species of persons if that species had been sufficiently similar in relevant respects to the actual species of human beings.

It is, furthermore, important to notice that my characterisation of personhood does not entail any form of Criterianism. This characterisation of personal identity is not a Criterianist account, since it does not, strictly speaking, establish any informative and non-circular strong criterion of identity for persons. As I have claimed, the marking feature of a Criterianist account of personal identity is that it entails that it is possible to specify non-circular and informative criterion of identity for persons where the criterion of identity also should entail a specification of a criterion of application for persons. But it is exactly the lack of an informative and non-circular specification of a criterion of application that makes my account a version of Non-Criterianism. Though I have characterised what it is to be a person, I have not established a criterion for what it is to be a person. The reason for this is that this characterisation of personhood is incomplete. It contains dots which indicate that an exhaustive specification of the functional features possessed by persons cannot be formulated.

This means that Animalism rightly points out that we are animals of a certain sort and that we exist as long as the animals exists. Contrary to Animalism, which is a Criterianist position, personal identity is on my view primitive in relation to animal identity, even though animal identity is a necessary condition for personal identity. Though we can give a weak characterisation of identity for persons of the form:

\[ x \text{ is the same person as } y \text{ iff } x \text{ and } y \text{ are persons, and } x \text{ is the same animal as } y, \]

we cannot conclude from this that personal identity is reducible to that of animal identity, since the characterisation of identity over time for persons relies upon a prior individuation of the objects as being persons and (i) that there exists no strict criterion of application for persons, and (ii) the possibility of individuating

---

200 See Chapter 3, section 3.4 for the notion of a strong criterion of identity.
myself as a person presupposes that the notion of ‘personal identity’ is applicable to myself. In this sense, remaining the same animal is a necessary and sufficient condition for me being the same person without constituting a non-circular criterion of personal identity.

Let me, from now on, call the account of person and personal identity which I have described above The Revised Animal Attribute View, or RAAV.

9.5 Objections to RAAV: material inadequacy of RAAV

Let me now consider some possible problems associated with the Revised Animal Attribute View.\textsuperscript{201}

a) Inconsistency with common-sense: Consider first the possibility that the RAAV seems to be inconsistent with common-sense verdicts of “the vegetable case” and the “foetus case”. Both cases reveal, it is claimed, that one cannot equate ‘person’ with ‘belonging to an animal kind K with an open-ended list of social and cognitive capacities’. The most plausible version of this argument is the Animalist version: according to this version individuals like you and I have been foetuses and we might end up in a vegetative state. But, in those circumstances it might be thought that we would not be persons, since we would then lack higher cognitive and social capacities. Therefore, individuals like you and I could be animals without being persons, but not vice versa.

Against this kind of problem I want to remind the reader about the case of Samantha in section 7.5. As I there claimed, our use of proper names and indexicals require, for a successful act of reference to occur, that the object referred to with a proper name or an indexical remains of the same sort as it originally was individuated as falling under. This entails that an individual lying in a vegetative state in a hospital which we are able to refer to with a proper name, such as ‘Samantha’ is still a person even though that individual no longer can refer to, or think of, itself as a person. That is, there is nothing wrong in thinking that something can be a person without actually implementing any higher order functions at all and that an individual in a vegetative state is a person in virtue of

\textsuperscript{201}I will not consider the alleged counterexamples of “body swap” and “teletransportation”. The reason for not considering them is that I have already established that personal identity over time requires, as a necessary condition, some form of physical continuity. But both cases of "body swap" and "teletransportations" are incompatible with the truth of that condition. Consequently, these two cases cannot constitute genuine possibilities in the life of persons.
the fact that it is an animal of a kind whose fully developed members have features normally associated with persons.

The same is also true of the foetus case. In thinking that you have been a foetus, you do not simply imply that the animal that you are has been a foetus. What you are implying is that you, the person who is doing the thinking that you have been a foetus, have been a foetus due to the fact that you individuate yourself as being essentially a person. Therefore, we do have personal identity even in cases such as the “foetus case” and the “vegetable case”.

b) Arguments from liberality and restrictiveness: Another possible problem for the Revised Animal Attribute view is that it is, at the same time, too liberal and too restrictive.\(^{202}\) I have claimed that a person must, by necessity, be an animal of some kind or other and that something is a person, not in virtue of satisfying certain cognitive capacities, but in virtue of belonging to a kind whose real essence explains the fact that members of the kind can possess certain cognitive faculties. But now, this means that we can have an individual human being who does not satisfy the conditions specified in the “capacity specification clause”, but we would still have to acknowledge that this individual is a person.\(^{203}\) But as Paul Snowdon says, it is only certain uses of person that fit this idea.\(^{204}\) It fits very well into the ordinary way of speaking about individuals lying, and dying in hospitals. In those circumstances we do use ‘person’ as a classificational term of individuals without any restrictions of mental endowment of these individuals. But this way of speaking does not fit everything we say about persons.\(^{205}\) For instance, some people are prepared to say that very young humans are not persons on the ground that they lack what is taken to be necessary mental capacities for personhood. Furthermore, the Revised Animal Attribute View is very restrictive in another sense. Suppose that a certain individual member of an animal kind that does not satisfy the “capacity specification clause” happens to be, as Snowdon puts it, “freakishly gifted”.\(^{206}\) Given the truth of the Revised Animal Attribute view, the only reason for not

\(^{202}\) Paul Snowdon, ‘Personal Identity and Brain Transplants’, in Human Beings, Supplement to Philosophy, 29, edited by David Cockburn, (Cambridge, Cambridge University Press, 1991), p. 102. This argument by Snowdon was originally designed against David Wiggins’ original version of the animal attribute view.

\(^{203}\) Snowdon, ‘Persons, Animals, and ourselves’, p. 102.

\(^{204}\) Ibid. p. 102.

\(^{205}\) Ibid. p. 102.

\(^{206}\) Ibid. p. 102.
classifying that individual as a person is that it happens to belong to the wrong animal species. But this seems to be counter-intuitive.

I do agree that these two consequences of RAAV are, to some extent, counter-intuitive. But with some reflection it is possible to reduce their air of counter-intuitiveness. Firstly, though it might seem to be incorrect to think that a foetus is a person under normal circumstances it is not equally counter-intuitive to think that a premature baby is a person. At least not from the point of view of the parents of the child and the care-takers of the child.

Secondly, regarding the possibility of a “freakishly gifted” animal the most reasonable thing to say is that we have reasons to suppose that its capacity for higher cognitive and social features is due to its real essence. That is, the fact that an animal can talk, think, engage in social norms and activities, etc., provides a good reason for assuming that the animal is constituted in such a way as enabling it to perform those things. If we did not take this attitude we would have to understand the “freakishly gifted” animal as one of the wonders of the world. It suddenly have the ability to engage in higher cognitive and social activities, but there is nothing in the animal itself that explains that it is able to do so. But I assume, this is nothing but magic, and magic is not something we have any reason to believe in. But if it is true that the “freakishly gifted” animal is constituted in such a way that enables it to perform all of those higher cognitive and social acts, then, the “freakishly gifted” animal belongs to an animal kind \(K\) with a real essence which would, if specified, give an account of the individual animals possession of higher cognitive and social capacities. This, in turn, entails that we have good reasons to assume that the animal, as well as the animals of the same kind as the “freakishly gifted” animal, are persons.

9.6 Objections to RAAV: persons and real essence

A more serious objection to the Revised Animal Attribute View is due to E. J. Lowe.\(^{207}\) Person is, I have claimed, not a natural kind on its own but a natural biological sort satisfying the condition that each individual member belongs to the sort Person in virtue of there being a natural biological kind of animals \(K\) such that (i) the individuals belongs to the kind \(K\), and (ii) the kind \(K\) has a real essence enabling its fully developed members to have a certain degree of cognitive complexity enabling them to think, reflect, perceive, remember, have emo-

---

tions and that the organism can take the same attitude towards other individuals of the world, etc.. Accordingly, one must accept that it is in virtue of being subject to certain empirical law-like principles that individual members of the sort person develop from their beginning of existence until their cessation. For instance, human beings are persons and it is in virtue of being subject to the inherent biological law-like principles governing the development of human beings that members of the kind Homo Sapiens can possess the higher cognitive capacities normally associated with being a person.

But characterising what it is to be a person in relation to a natural kind with a real essence in this way seems to indicate that person cannot be a biological natural sort, since different individuals persons could persist in virtue of having different real essences. Given the reasonable assumption that every individual animal persists in virtue of being subject to law-like principles distinctive of the sort it belongs to, then it seems to be the case that person, if it is a biological sort, ought to have law-like principles determining the possible development of individual persons. But assume for the moment that there exists a hitherto unknown kind of amphibian, bolgs. Assume further that bolgs satisfy the normally associated features of personhood. They are as strong cases of persons as human beings are. But Lowe's point is now that bolgs, if they are persons, cannot be subject to the same law-like principles as human beings. For instance, since they are amphibians we can assume that bolgs survive passing through having gills and a tail to having lungs and legs. But that is not an admissible alteration for a human being. The problem is that person, as a biological sort with its distinctive laws of development, either permit the change from having gills and a tail to having lungs and legs or they do not permit it. If they do, then it follows, absurdly, that an individual human being can survive the change qua person but cannot survive it qua member of Homo Sapiens. If they do not, then if follows, equally absurdly, that an individual bolg cannot survive the change qua person but can survive it qua bolg.

A possible reply to this argument is to claim that there is no real essence for person as such, since the real essence of an individual person would be the essence of the natural kind to which the person belongs. This is the line taken by Wiggins. According to this line of thought the world does not consist of persons simpliciter, but human-persons, dolphin-persons, bolg-persons, etc..

208 Lowe, Subjects of experience, p. 16.
209 Ibid. pp. 16-20.
210 Ibid. p. 20.
211 Wiggins, Sameness and Substance, p. 172.
Though this is a possible reply, it is not plausible on the ground that it is *person*, not human-person, that is the primitive concept by which we individuate ourselves. As I have claimed, knowing that one has a biological body, that one is a member of the biological kind *Homo Sapiens*, is secondary to knowing that one is a member of the sort *Person*. *Person*, then, is the primitive concept of our conceptual scheme.

But this does not imply that *person* is a non-biological concept. According to my theory, *person* is a biological concept in the sense that something is a person in virtue of belonging to a biological kind of the right sort. Given this position it is, I think, reasonable to assume the empirical hypothesis that there exists a uniform explanation of why individuals belonging to some biological kinds are persons while individuals belonging to other kinds are not. This uniform explanation, which might be knowable or not would (i) link mental phenomena to the neurophysiological structures which cause them, and (ii) determine what kind of biological complexity and organisation an animal must possess in order to be sufficiently complex to cause mental capacities normally associated with persons. This, though, does not imply that we have reasons to suppose only human beings to be persons. Other kinds of animals may also be persons, but if they are persons, then they are so in virtue of having a biological set-up resembling that of human beings in relevant respects. A fact explains why these kinds of animals typically have the capacities characteristic of persons. Thus, I think it is reasonable to assume that there exists a biological real essence (a set of properties) which is necessary and sufficient for a biological kind to be of the person sort.

But this account of the real essence of persons seems to give rise to a problem. Suppose, as I have claimed, that there exists a real essence *E* for person-hood, such that members of an animal kind *K* are persons in virtue of the fact that individual members of the kind have the real essence *E*. Is it not possible, contrary to my account, that one could give the following strong criterion of identity for persons?

\[ x \text{ is the same person as } y \text{ iff (i) } x \text{ and } y \text{ both belong to an animal kind } K \text{ whose members share the essence } E, \text{ and (ii) } x \text{ is the same animal as } y. \]

I do not think so! Even if there exists a biological essence which, together with a theory of mind and body, accounts for the fact that animals of some kinds *K* are persons and other animals are not, we should not conclude that it is possible to specify a strong criterion of identity for persons through the specification of the essence of persons. There is no plausibility in assuming that the real essence characteristic of persons will be specifiable in biological terms in one single formula. The essential properties possessed by persons may, if they are
specified, form an infinite, rather than a finite, set of true statements of necessary and sufficient conditions for personhood. In this sense, the essence of persons is analogous to a complete description of a state of the world. It is plausible to assume that a particular state of the world cannot be given a specification in form of one single formula, because of the highly complex character of that state of the world, but it is nevertheless plausible to think that there exists a description and an account of that state. In other words, it is plausible to assume that there exists a fact which would, if specifiable, give an account of personhood of the members of a biological kind K, while at the same time deny that this fact can be specified as a finite criterion of personhood.

9.7 Objections to RAAV: indeterminacy and transplantations

a) The indeterminacy thesis: It seems to be the case that questions of personal identity can be indeterminate because the life of an individual animal consists in the gradually continuous development of structurally united physical parts which preserves the life-process of the animal. Though the life-process itself might be an all or nothing fact, the continuity of the physical parts of the animal is a matter of degree. For instance, loosing certain biological parts such as an arm, a leg, a kidney or even a part of the brain does not imply the cessation of the animal. And as some Criterianists claim, since the essential continuities involved in cases of personal identity can hold to different degrees, it is also reasonable to assume that also personal identity is a matter of degree. Hence, some questions concerning personal identity can be indeterminate in truth-value. But this seems to be in conflict with my claim that personal identity is primitive, since the primitiveness of personal identity seems to entail the determinateness of identity, i.e., that questions regarding personal identity always have a yes- or no-answer.

Though it is a commonly known fact that animals develop gradually through the development of biological parts and that individual animals, as a consequence of this fact, might loose certain fundamental biological parts, I do not think this implies that personal identity is indeterminate. Firstly, since there is no doubt that the life-processes of the individual animal continue to exist after the loss of the part in question one should not think the degree of continuity of biological parts of the animal implies that personal identity is indeterminate. If an

---

animal is a person, then as long as the animal continues to be an animal of the sort S with a determinate identity it also continues to be a person with a determinate identity.

Secondly, there is an argument pointing in favour of upholding the view that, even though it is true that the essential continuity of an object can vary, the object's identity, in and by itself, is an all-or-nothing matter. This argument tries to establish that it is a metaphilosophical principle of identity that it is always associated with a determinate truth-value, such that questions of identity of objects can always be answered by a straightforward yes or no.213

Following Gareth Evans it is possible to show that, given our ordinary understanding of 'identity' as an equivalence-relation that necessarily is in accordance with the Indiscernibility of Identicals, then one cannot assume that the identity of objects can be indeterminate.214 If it is true that identity is indeterminate, then there must be a case such that;

(i) \( a=b \) is indeterminate.

But since nobody is prepared to accept that identity is completely indeterminate, everybody is prepared to accept that,

(ii) \( a=a \) is determinate.

But now, from the truth of the Indeterminacy of Identicals, it follows that premise (i) is false, since \( a \) has a property which \( b \) lacks; the property of \( \text{being determinately identical with } a \). Thus, we have, then, one reason to think that 'identity' is an all-or-nothing notion, and claim that there is no such thing as indeterminacy of identity. But from the fact that 'identity' does not admit of indeterminacy, I draw the conclusion that a person's identity, conceived as the identity of an individual object, cannot be indeterminate.

But in another, and more peculiar sense, it is true that personal identity can be indeterminate. Evolution makes it plausible to suppose that there exists, has existed, or will exist a large amount of biological organisms not easily classifiable in our zoological system. They are intermediary biological organisms in the

213 By 'straightforward' I do not imply that the determining of questions of personal identity is in any sense easy. For all we know, many difficult cases will arise. But all of these cases rest upon problems due to our ability to determine an answer to a question of personal identity, not upon the nature of personal identity itself. It is, then, 'straightforward' in the sense that there always exists an answer to the question and that the answer is always either yes or no that I intend by using the notion 'straightforward'.

sense that we cannot determine whether they belong to one species rather than another. Since I have claimed that something is a person if it belongs to an animal kind whose real essence is such that it explains the possession of certain higher cognitive and social capacities of fully evolved members of the kind it may also be the case that it is indeterminate whether a certain animal is a person. It is, to say the least, plausible to assume that there have been certain kinds of animals that possess some of the abilities listed in the Revised Animal Attribute View, about whom we nevertheless would be uncertain concerning the question of whether they were persons or not.

Imagine the existence of such an animal $a$ and imagine that it is indeterminate whether $a$ belongs to the species $S$ whose individuals are not persons or whether it belongs to the species $S^*$ whose individuals are persons. Since it is indeterminate whether $a$ is an $S^*$ it is also indeterminate whether $S^*$-identity obtains in this case. Consequently, since personal identity obtains in such a case only if $S^*$-identity obtains in such a case it is also indeterminate whether personal identity obtains in the case.

It should be pointed out that this form of indeterminacy of personal identity is something one has to accept if one is prepared to subscribe to a view in which biological species has a central role for person and personal identity. But, is this really a form of indeterminacy of identity? On the one hand, it is not reasonable to think it is a genuine case of indeterminacy, since it rests upon the assumption that it is meaningful to attribute personal identity to an object about which we do not know whether it is a person or not. But this assumption is not plausible. As I have claimed, individuation is sortal dependent: only if the object has been thought of as being this or that sort of object is it meaningful to claim that the object has been individuated. A consequence of this is that it is only meaningful to ask a question of identity of an object after it has been sortally individuated, since a question of identity of an object relies upon tracing the object under some sort $S$. But this is exactly what is missing in this alleged case of indeterminacy of personal identity. It has not been determined whether the object is a person or not. Hence, we do not know whether personal identity obtains or not. But this does not imply that the identity of the individual object is indeterminate. What this implies is that questions of

215 An analogy would be the following. Imagine the discovery of a new celestial configuration in the universe. Scientists debate whether the newly found entity is a star or a planet. Since it is indeterminate whether the entity is a star or a planet it would also be indeterminate whether star-identity obtains. But to claim that the entity's, star or planet, identity is indeterminate because of this seems out of place. Either it is a star and star-identity obtains or it is a planet and planet-identity obtains. It might be thought that this example is different from the case of person, since something is a person in virtue of being an animal of
identity are only meaningful after sortal individuation and that personal identity should not be understood as indeterminate. Given that the object is individuated as a person no such indeterminacy remains.

Furthermore imagining, for the sake of argument, that it is meaningful to attribute personal identity to an object about which we do not know whether it is a person implies that no concrete substantial object ever could be determinate, since every concrete object can be said to pass through a certain stage in which it is unclear whether the object exists or not and every sort of object can be said to pass through a stage in which it is unclear whether there is anything of the sort S or not. For instance, when does a particular house begin to exist? Or a dog? A certain structural unity is necessary, but how much? Furthermore, at what precise moment does a sort begin to exist? One has only to think about the different hominids in the evolutionary history of our own species. Neither Australopithecus nor Homo Habilis were human beings, but what about the Neanderthals? What we can say is that there has been a slow and gradual evolution of animals which has resulted in the sort human being. We cannot say when the first human being began to exist. In this sense, then, it is indeterminate when an animal begins to exist and when a sort begins to exist and personal identity is, at least, as determinate as one might expect objects to be in an evolving universe.

b) Brain-transplantation:216 I have claimed that a person continues to persist as long as the animal persists. But remember “the Brownson-case” where the brain of Brown was put in the body of Robinson. The intuitive response to that situation was that the person Brown would survive the operation, but now having Robinson’s body. If we interpret ‘body’ as standing for the animal, it seems that this is in conflict with my claim that personal identity follows animal identity. Schematically a brain-transplantation case, such as Brownson, can be structured as follows:

(i) it is possible to perform brain-transplantations,

(ii) it is the brain that sustains higher cognitive and social capacities,

(iii) in performing a brain-transplantation the animal does not follow the brain,

(iv) in performing a brain-transplantation the person follows the brain.

a kind S and it was indeterminate whether the object was of that kind S. If this is a problem, simply imagine the existence of a planet system around the newly found heavenly entity and suppose it to be the sun of that solar system.

216 I am very much in debt to Paul Snowdon for both the way in which the case of brain-transplantations are presented, and in the responses I take towards cases of brain-transplantation.
Concl. Personal identity is distinct from animal identity

In brain-transplantation cases it seems to be the case that we have *sameness of person* without necessarily having *sameness of animal*.

What I would like to question in this form of argument is premise (iv). It is, I think, an implicit assumption of this form of argumentation that the mental content caused by the brain is internally self-sufficient. If we take my brain and locate it in a new body, the mental content caused by that brain will not change as a result of its new environment. But why should we accept that? What I presently experience, think, feel, etc. depends to a large extent upon the fact that a certain brain is located in a certain living body that is interacting with the natural and social environment in a certain way. For instance, the visual experiences that I have of the world are from a certain point of view which depends upon how tall I am, and my self-conception depends to a large extent upon the physical capacities of my biological body.

What this entails, I think, is that one should understand the brain as a cognitive organ of an animal. It is a very important organ, though, since it is the organ we use in structuring our behaviour in relation to the world. But it is nevertheless, an organ of an animal. Since the brain is an organ, it does seem reasonable to assume that a transplantation of a brain, as a transplantation of other biological organs, requires a certain period of adaptation to the new environment. This adaptation to its new environment would consist in the mental content caused by the brain being quite different from the earlier mental content, because of the differences between the bodies that the brain is located in. But if the mental content caused by a brain that is transplanted adapts in this way to its environment, it seems that there are no real reasons to suppose that the brain continues to sustain one single person. This means that it is reasonable to assume that a successful transplantation of a brain to a new environment does not entail that a person goes where the brain goes.

However, there are no strong reasons to assume that it is reasonable to assume, as for instance Paul Snowdon does, to think that the person sustained by the brain in the new environment is identical with the animal that is the receiver of the brain. The reason for this is that the animal has to adapt to the new brain, *qua*, cognitive organ. Firstly, it is reasonable to assume that there exists certain differences between brains which influence the way in which an animal acts and thinks. In this way, the body must also adapt to the deficiencies and the strength of the new brain. Secondly, it is unreasonable to assume that all the

---

217 This view has been defended by Paul Snowdon both in discussions and seminars.
mental content caused by a brain need to be restructured after a successful transplantation. For instance, a certain production of hormones might be preserved after the transplantation such that the receiver of the brain acts and think differently than before the operation. What this entails is a view of brain-transplantation as the creation of a new animal, as well as a new person. That is, it is more reasonable, since the mental content differs as a consequence of the transplantation, that a new person and a new animal begins to exist after the successful transplantation has been performed. But this entails that brain-transplantation cases cannot constitute a counter-example to my account of personal identity.

c) *Head-transplants:* A further objection against my account of what it is to be a person is the possibility of head-transplants. Suppose you are visiting a friend at the hospital. Your friend is lying in his bed with a blanket over his body. All you see is his head. The two of you talk about how he feels, what he will do when he leaves the hospital, that his children are doing fine, etc. All in all, then, you have no trouble in thinking that the individual lying in the hospital-bed is your friend. Imagine now that as you are talking, the blanket suddenly falls to the floor. When you look at your friend you realize that there is no longer any body attached to his head. All you see are cables and wires running directly from his head to different machines at the far end of the bed. These machines sustain the person who formerly was sustained by biological processes of his body. But since it is reasonable to assume, the argument continues, that a single head does not make an animal, there is no longer an animal lying in the hospital-bed. But this means that my account of what it is to be a person is incorrect, since I have claimed that remaining an animal of a certain kind is a necessary condition for being a person.

How should one respond to this kind of case? Let me first say that if this situation occurred, then I do think the correct immediate response would be bewilderment. A point at which we simply do not know what to say. But a little reflection will reveal, I think, that there are two possible responses one could take. On the one hand, it is questionable whether this kind of thought-experiment really establishes a genuine possibility. The whole situation is underdescribed. We are told nothing about how the machines are supposed to be able to sustain the person. Also, in virtue of being highly under described, the possibility of head-transplants is really a non-starter for elucidating what it is to be a person.

---

218 By "head-transplantation" I also mean the pure preservation of a head, or a "body-amputation".
Secondly, why should one accept that the head lying in the bed is not an animal? For all we know, animals can lose parts of their bodies without ceasing to exist as long as one and the same life-process is preserved. How much, exactly, one might lose is a matter of empirical investigation. But given that in the head-transplant case we have personal identity, why should we not think it is animal identity? It is one and the same life that continues to persist. The only difference is that it is sustained by non-standard external life-supporting devices. But this cannot really make a difference, since many animals are sustained by non-standard external life-supporting devices (for instance, during certain operations). We have no reasons to think the head in the bed to be anything other than an animal. The head in the bed is an animal reduced to (probably) a minimum of what is necessary for remaining an animal. But, this means that head-transplants simply cannot be considered as counter-examples to my account of persons and personal identity. Persons are animals of a kind whose real essence is such that, if specified, it would give an account of the possession of certain higher cognitive and social faculties of fully developed individuals of the kind, and if personal identity obtains in a certain situation, then also animal identity obtains in the same situation.

Let me end this discussion with a further consideration which does point in favour of the claim that it is essential, i.e. a necessary condition, for being a person that one is a biological animal. What I am thinking of is the fact that we experience great stress and fear in knowing that the life-supporting biological process of one self, or some person one knows, is threatened for some reason or other, and that under normal circumstances we are prepared to go to great effort, even through terrifying pain, with the aim of preventing the cessation of these biological processes. One has only to think of the fear of a parent whose small child has swallowed a coin, and where the coin has become stuck in his throat so the child cannot breathe, or the enormous pain involved in cytotoxic treatment of leukaemia and other forms of cancer. In the first case, the parents are fearing that the stuck coin will end the child’s biological processes, and in the other case, the cytotoxic treatment, and the pain involved, is something a person is prepared to go through in the aim of prolonging the biological life-processes of that person’s body.

Imagine that your son or daughter is diagnosed as having leukaemia and that he or she is going through cytotoxic treatment. As a parent you experience enormous distress seeing your child going through all the pain and horror associated

---

219 The same conclusion has been reached by Peter van Inwagen. (See his *Material Beings*, (London: Cornell University Press, 1990), pp. 177-181.)
with the cytotoxic treatment. But you are, of course, also experiencing distress and anxiety because you are afraid that the chemotherapy will be unsuccessful and not be sufficient to save the life of your child. Suppose, now, that someone comes up to you and claims that it is possible to save your son or daughter without forcing him or her to go through the horrors of the cytotoxic treatment. “What we do”, the person continues, “is to give your child an anaesthetic, and then put him or her in a transformation-process where all the biological structures of the child’s body are slowly and gradually replaced by inorganic, bionic material. Since the child has been given an anaesthetic the whole process is carried out without any experience of pain by your child. Since this process is carried out slowly and gradually, all the mental functions of the child are preserved, even though the biological foundation of these mental functions are changed. So, for instance, after the transformation-process is done, the child will recognise you as her or his father, remember all her or his friends, where she or he lives and so on. “Of course”, the person continues, “she or he will no longer be an animal, a biological organism, but nothing important will have changed. The only thing you have done by letting us treat your child is to save your child from going through a horrifying and painful experience such as the classical chemotherapy.”

A plausible response to make if you were presented with such an option is, I think, to question the sanity of the person who offers you the “alternative treatment”. Your child is not only the existence of a certain physical body with a certain distinctive psychology, it is a living biological person of flesh and blood. And to think, as the person with the “alternative treatment” must be thinking, that the flesh and blood of your child has nothing to do with your responsibilities, caring and loving of that child is just insane. From your point of view, what the person is offering you is not so much the rescuing of your child from experiencing pain and horror, but the extinction of your child all together. That is, the extinction of the biological person that your child is.

9.8 Concluding remarks

I have in this essay tried to defend a particular view of personal identity in which the basic meaning of ‘person’ and ‘personal identity’ is determined. It is an account which has as its objective the determination of what it is to be a person from the perspective of how we basically use the terms ‘person’ and ‘personal identity’. I defended a thesis in which the concept person is understood
Persons—Their Identity and Individuation

as a basic sortal concept which picks out a natural biological sort of enduring persons. A person is, according to this view, an object of the world which remains the same over time in virtue of being numerically the same at every moment of its existence and which belongs to the sort Person in virtue of having certain essential properties.

The claim that person is a basic sortal involves five different but interrelated claims. The first is that the sortal concept person is a primitive sortal concept in our conceptual scheme in the sense that we, by necessity, must individuate ourselves as being of the sort person. The second claim entailed in regarding person to be a basic sortal is that it is a substance sortal in that an individual object falling under 'person' ceases to exist if he no longer falls under 'person'. According to this point person is the substance sortal which picks out what we fundamentally are, and hence that we essentially are persons in the sense that we cannot continue to exist without continuing to exist as being persons.

Both of these points have been questioned by Animalists. According to Animalists, the substance sortal 'animal' or 'Homo Sapiens' better picks out what we fundamentally are; and they claim that it is possible to individuate ourselves as belonging to that sort without individuating ourselves as persons. It is 'animal' or 'Homo Sapiens' which picks out our essential nature in the sense that we cannot cease to be animals or human beings without thereby ceasing to exist. 'Person' is only a phased sortal which singles out a contingent psychological feature of individuals of the kind Homo Sapiens, i.e., it is not necessary for the preservation of the individual object that it continues to satisfy these psychological features. The reasons for not accepting the Animalist position are (i) the determination of ourselves as belonging to the kind man presupposes an irreducible first-person perspective which is intelligible if and only if the concept person is understood as the primitive concept by which objects like ourselves individuate themselves and each other, and (ii) that our very use of certain proper names and indexical terms such as 'I', 'she' 'he' and 'you' are person-inducing in that a successful reference to an object with such a proper name or an indexical relies upon the fact that the object is individuated as a person.

Thirdly, claiming that person is a basic sortal entails that an individual person is an object who persists in virtue of being wholly present at each time that that individual exists. Persons are, in this sense, three-dimensional objects in that they endure over time. In this, as well as throughout the whole essay, I have relied upon an understanding of 'identity' as a primitive notion in our scheme of thought and that it always stands for numerical identity. In this connection, I argue against the Perdurance view of persons which maintains that persons are
four-dimensional objects extended in both space and time and that persons persist in virtue of having distinct temporal parts at distinct moments.

The fourth point in claiming that person is a basic sortal is that personal identity is understood as something primitive in our conceptual scheme, since we cannot provide a criterion of identity over time for persons. We simply cannot specify non-circular and informative necessary and sufficient conditions for personal identity. The argument for upholding this claim is the inadequacy of hitherto presented accounts of the criterion of identity over time for persons. In this sense, the argument is a form of inference to the best explanation. Since the best available accounts of personal identity or our identity are insufficient to provide us with a criterion it is reasonable to conclude that no criterion of personal identity can be found. Claiming this makes my view Non-Criterianist in that it upholds: (i) the unspecifiability thesis, and (ii) the primitiveness thesis. The first thesis maintains that personal identity cannot be non-circularly specified in terms of necessary and sufficient conditions, while the second thesis is that personal identity over time is primitive in relation to the empirical continuities that might underlie it. Contrasted to such a Non-Criterianist account is the Criterianist position which consists in accepting: (i) the specifiability thesis, (ii) the complexity thesis, (iii) the derivativeness thesis and (iv) the essentiality thesis. According to Criterianism it is possible to specify, in a non-circular way, necessary and sufficient conditions for personal identity over time, or identity over time for objects belonging to the same kind as us, since the fact of personal identity or our identity depends upon some more basic facts which must be semantically-cum-metaphysically understood if personal identity or our identity should be understood. The last thesis of Criterianism concerns the fact that the specified necessary and sufficient conditions should capture what essentially is involved in cases of personal identity, or the identity over time for objects of the same kind as us. I examine three different views which all agree in the idea that it is possible to specify personal identity in more fundamental terms, i.e., Psychological Criterianism, Physical Criterianism and Animalism.

Psychological Criterianism fails because it cannot be formulated without presupposing personal identity over time. I argue that both Locke’s original theory which exclusively focused upon a person’s memories, and more recent versions of Psychological Criterianism are viciously circular. This is even true of accounts which do not rely upon, the previous awareness condition i.e., accounts which take it to be possible that there are quasi-cognitive faculties providing persons with the same kind of knowledge ordinarily provided by normal cognitive faculties. According to me, all of these Criterianist accounts are viciously
circular, since (i) mental features, such as memories, are causally dependent and it is not possible to discern whether a particular mental feature of a person is a genuine mental feature or not without presupposing personal identity. And (ii) the notion of 'quasi-psychological faculties' is conceptually tied to personal identity in the sense that we cannot determine what it is for a person, from the first-person point of view, to have an apparent mental feature without invoking personal identity as a part of the content of the mental feature in the form of a belief. For instance, it is not possible to determine, from a first-person point of view, whether a memory impression is a quasi-memory or a genuine memory without presupposing personal identity over time, since the apparent remembering of an event E by a person P is accompanied by an apparent belief held by P which makes the content of the apparent memory 'I E-ed', or 'I am the same person as the one who E-ed'. A further point argued for is that it is only possible to individuate mental states through a prior individuation of the person who is the "bearer" of the mental states. It is contrary to what the advocates of Psychological Criterianism think is the case, the fact of personal identity which explains the continuity of a person's cognitive capacities and mental content, and not the other way around.

The reasons for not accepting Physical Criterianism, both the criterion focusing upon the brain and the criterion focusing upon the body, is that (i) neither version establishes a sufficient condition for personal identity over time, and (ii) having certain physical parts is not sufficient for determining that an object is a person. Both versions of Physical Criterianism hold that it is the spatio-temporal continuity of a material entity, the brain or the body, which constitutes or grounds personal identity over time. But it is possible to show that spatio-temporal continuity cannot guarantee identity, since it is possible that two individual entities can be spatio-temporally continuous with an original entity without thereby being identical with each other. That is, spatio-temporal continuity cannot ground the transitivity of identity. Now, the attempt to solve the problem of splitting persons through introducing a principle saying that personal identity consists in spatio-temporal continuity as long as the spatio-temporal chain is of a "non-branching" kind is not sound since it makes identity an external relation. But there are strong reasons to suppose that a metaphilosophical principle of 'identity' is that the holding of identity between x and y only depends upon the internal relation between x and y, and not upon the relation of x and y to another object z. It is one of those presupposed principles structuring questions about identity, since it is part of the conceptual framework in which questions of identity are asked and answered. Another metaphilosophical principle is that
identity is determinate, since this idea rests upon the principle of bivalence. That is, it is part of the framework that it should always be either true or false that a certain later person is identical with a previously existing person. Disregarding theories relying upon the non-branching clause, then, a purely physical criterion cannot establish a principle by which we might distinguish persons from non-persons, since continuity of spatial parts, in and of itself, is something which persons and non-persons share, hence, Physical Criterianism cannot help us establish what it is to be a person.

Animalists hold that the central question in the area of personal identity is the question about our identity: what are we, individuals like you and I, essentially and what does our identity over time consist in? Their answer is that we are human animals and that our identity over time consists in being the same animal, or the same biological organism. This is, I think, the best available Criterianist version, and to a great extent I agree with Animalists. But Animalists have missed the fact that we must, by necessity, individuate ourselves as belonging to the substance sort person and that this fact makes Animalism untenable as an account of the necessary and sufficient conditions for our identity over time.

The fifth claim contained in the idea of person being a basic sortal is that 'person' is understood as referring to a substantial sort, i.e., a real and natural sort of individuals existing in the world. This claim is an acceptance of realism concerning the sort Person as well as individual persons. The sort Person consists of individual persons who have, in and of themselves, a certain unity and structure which determines the growth, development, alterations and cessation of persons. The sort Person is a sort in the world which consists of individuals with a certain essence, and the sortal concept person would, if possible to specify, give an explanatory account of the essential properties of persons as persons.

Though person is a basic sortal concept, I claimed it to be possible to specify some conditions for personhood and personal identity. I argue that something is a person only if it is a biological organism or animal which belongs to a certain kind whose real essence is such that it would, if specified, give an account of the possession of certain higher cognitive and social capacities of fully developed members of the kind. This account accepts the Animalist identification of ourselves as human beings, while retaining the insight of Locke and the psychological tradition that having psychological capacities are important features of being a person. But contrary to Animalism, we are essentially persons, and contrary to the Lockean tradition we are not essentially psychological beings. The argument for the first part of this characterisation of what it is to be a person is (i) the fact
that predicates attributing mental states and consciousness, P-predicates, are matter-invoking in that they can be truly applicable of an object only if the object also has certain physical states, (ii) the fact that these physical states are attributed to an object via M-predicates presupposing bodily characteristics, and (iii) that the meaning of our concept *person* is determined in relation to what we know to be persons, i.e., the actual persons of the world which are biologically constituted in the same relevant respects as individuals like you and I. The argument for the second part of the characterisation of what it is to be a person is (i) the externalistic account of meaning for *person*, and (ii) the necessity of individuating ourselves as persons.

I further claim that *person* should be understood as a biological qualification of *animal*. *Person* is, on my view, a biological concept, since the concept *person* has a necessary biological conceptual constituent.

Claiming that persons are biological organisms distinguishes my account of persons from Psychological Non-Criterianists: Cartesianism, The Subjective View and Psychological Substantialism. What these accounts have in common is a more or less exclusive first-person perspective on why Criterianism fails. This first-person perspective leads Psychological Non-Criterianists to maintain that persons are, primarily, psychological entities. Against Cartesianism and The Subjective View I argue that (i) their account rests upon three Cartesian mistakes, and that (ii) the application of mental predicates to a person is ontologically dependent upon the correct attribution of material predicates to the person. The three Cartesian mistakes are (i) the *fallacy argument* which consists in thinking that the fact that my knowledge of myself does not depend upon what I am unaware of shows that what I am unaware of is not what I am essentially, (ii) the *mistake from the ways of thinking of ourselves*, i.e., the view that first-person thinking can establish what are non-essential and what are essential properties of persons, and (iii) the *imaginability mistake* which consists in confusing logical possibility with metaphysical possibility. The second point is that all P-predicates applicable to a person are matter-involving in the sense that if certain M-predicates were not true of the person, neither would the P-predicates be true of him. Against Psychological Substantialism, I argue that it is not reasonable to think of persons as psychological substances, and hence *person* is a psychological substantial kind. The reasons for this are (i) there is, as Animalists point out, a strong intuition in favour of identifying oneself and other persons with a foetus and an individual in a vegetative state, and (ii) that we have reasons to suppose that animals can have the same higher cognitive capacities as persons and since the higher cognitive capacities of animals should be biologically explained it
seems reasonable to suppose also that persons' higher cognitive capacities should be biologically explained.

I also claim that persons have an essence *qua* persons. Though different kinds of animals might have different essences, there might also be a certain degree of similarity of essence between animals of different kinds, and particularly it might be the case that animals with the same type of cognitive abilities have similar essences. The reason for assuming this is that it is reasonable to assume that mental activity is based in biological laws and that the same biological laws structure the same, or similar observable phenomena. This means that persons have a real essence *qua* persons. But by claiming that persons have a real essence I do not imply that it is possible to give a non-circular specification of what it is to be a person in terms of the biological essence. It is still the fact that *person* is the primitive concept by which we individuate ourselves. We are, first and foremost, individuals of the biological sort *Person.*
BIBLIOGRAPHY


Evans, Gareth, ‘Can There Be Vague Objects?’, Analysis, 38, 208.


Kripke, Saul, ‘Identity over time’ Unpublished manuscript.


Nelson, John C., *Some comments on the identity conditions for material objects*, A Dissertation submitted to the Faculty of the Division of the Humanities, Department of Philosophy, The University of Chicago, (Chicago, 1970).


Parfit, Derek, 'Personal Identity', in *Philosophical Review*, vol. 80, 1971.


INDEX

alternative biography 161-162
animal attribute view 138, 199
Animalism 18, 71, 137-156, 168, 198
basic sortal 16, 19, 177, 187-219
biological sort 16-20, 187-219
Body criterion 18, 122-136
Brain criterion 18, 110-113
Cartesianism 18, 157-160, 164-171
circularity objection 93-97, 102-105
complexity thesis 70, 72-74
continuant 7, 61-66 175
Criterianism 7, 17-8, 67-83
criterion of identity 34-39, 44-50, 215
derivativeness thesis 69, 74-75
embodiment 126-133, 167-171
endurance view 17, 42, 54, 60-66, 153
enduring 7, 16, 41, 54, 190, 214
essentiality thesis 70, 75
fission 32-33, 134-135
foetus case 144-146, 183, 201
four-dimensional object (language of)
6, 42, 54-60, 120, 190
fusion 32
human being 18, 137-155, 184, 214
identity 4-9, 21-33, 214
determinateness 26, 206-209
extrinsicality 117-122
indeterminacy 58
intrinsicality of 25, 117-122
numerical 6-9, 16, 21-33, 153, 214
qualitative 21, 26-27, 119
relative 36-39
indeterminacy thesis 77-79, 206-209
Indiscernibility of Identicals 22-26
individuation 17, 41-66
M-predicate 126-133, 170-171, 185-186, 188, 218
M-properties 170, 185-186
multiple personality 172, 178-186
natural kind 46-47, 191-219
Non-Criterianism 7, 17-20, 67, 84-86
ontological priority 52-66
P-predicate 1, 92, 127-133, 151-153, 168, 185-186, 218
P-properties 185-186
perdurance view 42, 54-60, 64-66
phased sortal term 41-52, 137-155
Physical Criterianism 18, 71, 109-136
Physical Realiser criterion 18, 122-136
previous awareness condition 93-99, 215
primary substance 61-66
primitiveness thesis 85, 157
Psychological Criterianism 18, 71, 87-108
Psychological Non-Criterianism 157-184
psychological substantial kind 175-184
Psychological Substantialism 18, 157, 171-186
quasi-memories 97-108
real essence 138-141, 184, 191-219
reductionist thesis 76, 79-83
reduplication 117-122, 133-135
Revised Animal Attribute View 19-20, 201-219
Revised Brain criterion 113-122
simple substance 175-177
sortal concept 41-66
sortal dependency of individuation 17, 41-66
sortal term 41-52
specifiability thesis 69, 70-72
split-brain 8, 107-108, 172, 178-181
The Subjective View 18, 157, 160-167
substance sortal term 19, 41-52, 188-192
substance-dualism 159-160
substantial kind 171-177, 183-184, 194
substantial sort 19, 187-219
thought experiment 10-16
three-dimensional object (language of) 7, 54, 60-66, 190
unspecifiability thesis 85, 157
5 Sigurd Fries. Svenska växtnamn i riksspråk och dialektt. 1975.
9 Sylvia Danielson. Samuel Columbus’ språkprogram i En Svensk Ordeskötsel. 1976.
27 **Algot Hellbom.** Ordbok över Njurundamålet. 1980.
29 **Veronica Bonebrake.** Historical Labial-Velar Changes in Germanic. Diss. 1979.
31 **Maria Bolander.** Predikativens funktion i svenskan. Diss. 1980.
32 **Konstkultur i Norrland.** En informationsskrift om forskning vid Institutionen för konstvetenskap. 1980.
33 **Internordisk språkförståelse.** Utg. av C.-C. Elert. 1981.
34 **Magnus von Platen.** Privatinformation i skolan. En undervisningshistorisk studie. 1981.
37 **Svenskans beskrivning 12.** Utg. av C.-C. Elert og S. Fries. 1981.
40 **Claes-Christian Elert.** Ljud och ord i svenskan. 2. 1981.
43 **Readings on Unbounded Dependencies in Scandinavian Languages.** Edited by Elisabet Engdahl and Eva Ejerhed. 1982.
49 **Nordsvenska.** Språkdrag i övre Norrlands tätorter. Utg. av C.-C. Elert och S. Fries. 1982.
52 **Tage Boström.** Okänd soldat och kända soldater. Beteenden, attityder och struktur i Väinö Linnas krigsroman. 1983.
54 **Birger Liljestrand.** Tal i prosa. Om svenska författares anföringsteknik. 1983.
71 Hugo Olsson. La concurrence entre il, ce et cela (ça) comme sujet d'expressions impersonnelles en français contemporain. 1985.


La culture dans la langue. I Söhrman (Red.) 1993.


Gunnar Persson & Mats Rydén (eds). Male and Female Terms in English. 1996.


133 **Clarence Larsson.** Identity Through the Other: Canadian Adventure Romance for Adolescents. Diss. 1996.


140 **Mats Rydén, Henryk Kardela, Johan Nordlander and Bengt Odenstedt.** From Runes to Romance. A Festschrift for Gunnar Persson on his Sixtieth Birthday. 1997.


143 **Roger Melin.** Persons - Their Identity and Individuation. Diss. 1998

---

**Försäljning och distribution:**

Swedish Science Press, Box 118, SE-751 04 Uppsala Sweden.
What is a person and what is it that makes a little child that we once met the same person as the adult that we meet much later in life? What does our own identity over time consist in?

This book is about the nature of persons and personal identity. It belongs to a tradition that maintains that in order to understand what it is to be a person we must clarify what personal identity consists in. The author divides the problem of personal identity into two problems (i) How do persons persist? and (ii) What facts, if any, does personal identity consist in?

Concerning the first question, the author argues that persons persist three-dimensionally (the endurance view), and not four-dimensionally (the perdurance view), on the ground that objects must always fall under some substance sortal concept $S$ (the sortal dependency of individuation), and that the concept person entails that objects falling under it are three-dimensional.

Concerning the second question, the author differentiates between Criterianism, the view that it is possible to specify a non-circular and informative criterion for personal identity, and Non-Criterianism, which denies that such a specification is possible. He argues against Criterianist accounts of personal identity on the ground that they are either (i) circular, (ii) violate the intrinsicality of identity or (iii) do not adequately represent what we are essentially. The author further criticises three Psychological Non-Criterianist accounts of personal identity on the ground that they wrongly assume that ‘person’ refers to mental entities.

Instead the author formulates the Revised Animal Attribute View where person is understood as a basic sortal concept which picks out a biological sort of enduring animals. In this, he claims that the real essence of a person is determined by the real essence of the kind of animal the person is, without thereby denying that persons have a real essence as persons.