

**PERSONAL
IDENTITY AND
SELF-
CONSCIOUSNESS**

Brian Garrett



London and New York

First published 1998 by Routledge
11 New Fetter Lane, London EC4P 4EE

This edition published in the Taylor & Francis e-Library, 2002.

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

© 1998 Brian Garrett

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalogue record for this book has been requested

ISBN 0-415-16573-3 (Print Edition)
ISBN 0-203-01566-5 Master e-book ISBN
ISBN 0-203-22346-2 (Glassbook Format)

THE PROBLEM AND ITS PLACE IN PHILOSOPHY

The problem of personal identity

This book is intended as an overview of issues in the philosophy of persons and personal identity. In the first five chapters, we will be concerned with questions dealing with the nature (or metaphysics) of persons and personal identity. In the sixth chapter, we address the question of whether the value or importance that we attach to persons and personal identity is justified. In the final two chapters, we shall assess the extent to which a proper understanding of the semantic (that is, meaning-related) and epistemic (that is, knowledge-related) features of first-person judgements—judgements of the form ‘I am F’—can shed light on the concept of self-consciousness. This concept is a key constituent of our concept of a person.

The concerns of this book are strictly philosophical. We are not concerned with issues of ‘personal identity’ as this phrase is colloquially understood (in terms of a person’s self-image or fundamental values and beliefs). Rather, we are concerned with personal identity in an abstract way, where what matters is not the particular characteristics that distinguish us, but those characteristics we all (or most of us) have in common.

Moreover, the word ‘identity’ should be taken to connote strict numerical identity, not mere qualitative identity (that is, exact similarity). The distinction between numerical and qualitative identity is crucial in what follows. We are not concerned with identity in the sense of qualitative identity or exact similarity, as when we talk of identical twins or identical billiard balls. Rather, we are concerned with identity in the sense of ‘numerical identity’. In this sense, twins are not the same, they are two different people. Throughout the book, ‘identity’ should always be understood in this second, numerical, sense, unless otherwise indicated.

One might wonder why there should be a problem about personal identity. Is the relation of personal identity not simply an instance of the relation of identity, and so defined by the formal properties of reflexivity ($(\forall x) (x=x)$) and congruence ($(\forall x)(\forall y)(x=y \rightarrow x$ and y share all their properties)? A relation R is the relation of identity just if R is reflexive and congruent. What more needs to be said?

The answer, fortunately, is that a lot more needs to be said. The formal properties of identity tell us absolutely nothing about why we are right to make many of the judgements of personal identity that we do make, both in ordinary cases and in more outlandish fantasy cases.

For example, suppose we rightly judge Moriarty to be the murderer. We can ask why this is true. Someone might respond: Moriarty is the murderer because Moriarty stands to the murderer in the relation of identity, defined as above. However, it would be a fallacy to think that the availability of such an unilluminating response implies that there are no non-trivial necessary and/or sufficient conditions ('criteria') for the truth of judgements of personal identity.¹ Such judgements are subject to material conditions of correctness, and the formal properties of identity can tell us nothing about those conditions.

We can think of the matter as follows. The sentence 'A is the same person as B' is equivalent to the sentence 'A is B, and A and B are persons'. The truth of such sentences is subject to two sets of constraints: the formal constraints of identity, and constraints that follow from what it is to be a person. The task of the first five chapters of this book is to elucidate these latter constraints. The methodology employed is unrepentantly *a priori*.

In the chapters that follow we will be concerned to answer the following questions:

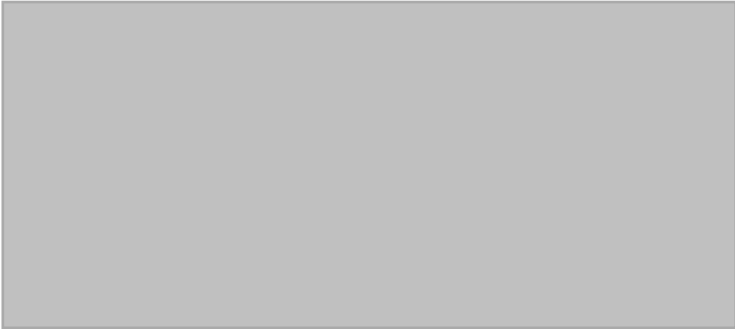
- What is a person? Spirit, animal, body, brain?
- What is it for the same person to persist through time? Can I survive the destruction of my body and brain? Can I survive the extirpation of my mental life?
- What does the possibility of fission show about the nature and importance of personal identity? The example of fission which will concern us is an imaginary case in which surgeons bisect my brain and transplant each hemisphere into its own body, resulting in the creation of two people, both of whom are psychologically very like me.
- Is personal identity an all-or-nothing matter? Or can it sometimes be *vague* or *indeterminate* whether a person at one time is the same

as some person at a later time? Can it be a vague matter whether I will exist tomorrow?

- Is the special importance we each assign to our own futures irrational? That is, is personal identity really the justifier of the ‘special’ concern which we have for ourselves in the future, or is the justifier some other relation which accompanies personal identity in the normal case?

For example, Derek Parfit thinks that the relation of *psychological continuity* is the justifier of the ‘special’ concern we have for ourselves in the future.² This relation is composed of a number of chains, or strands, of interlocking direct psychological connections, such as those which hold between an experience-memory and the experience-remembered, or between an intention and the action which manifests it, or the chain consisting of the retention of beliefs, desires, memories, character, etc., over time. The relation of psychological continuity is not the same as the relation of personal identity, as the possibility of the fission and fusion of persons makes clear.

-
-
-



What is a person?

In asking a question of the form ‘What is an F?’, we are asking a question in ontology. It is a question about the nature of Fs, not a question about the meaning of ‘F’ or the concept of F-ness. However, a question of the form ‘What is an F?’ is often ambiguous. It can mean: ‘What conditions does something have to satisfy in order to be an F?’ (call this the satisfaction question). Alternatively, it can mean ‘Of what kind of stuff (animal, vegetable, mineral, etc.) are Fs composed?’ (call this the nature question).

Thus, the question ‘What is a table?’ can be disambiguated in either of these two ways. In the first way, taken as a satisfaction question, the appropriate answer would be ‘A table is an object, typically man-made, and typically having four legs, which is used for putting coffee cups on,

THE PROBLEM AND ITS PLACE IN PHILOSOPHY

working on, eating off, etc.’. In the second way, taken as a nature question, the best answer would be ‘A table is an artefact which is not made out of any one kind of stuff—tables can be made out of no end of material (wood, aluminium, plastic, gold, ice, etc.)’.

Notice that in the case of ‘What is a table?’ the two answers are independent of each other. In particular, the answer to the satisfaction question does not determine any particular answer to the nature question. The knowledge that something is a table (in the satisfaction sense) does not allow us to form any expectations about its composition. This is not so, however, in the case of a question such as, for example, ‘What is a tree?’. The answer to this question, understood as a satisfaction question, cannot be separated from the answer to the question, understood as a nature question. Trees are necessarily made of wood, and a full answer to the satisfaction question will have to make reference to this fact.

Consider now the question ‘What is a person?’. This can be understood either as a satisfaction question or as a nature question. In this and subsequent chapters, we shall be concerned to answer both the satisfaction and nature questions, and to assess the relation between them. We should then be in a position to determine whether an answer to the question ‘What is a person?’, understood as a satisfaction question, is independent of the answer to that question, understood as a nature question.

Some philosophers believe that the best answer to the satisfaction question is not independent of the best answer to the nature question. According to the animalist of Chapter 2, for example, the best answer to the satisfaction question will have to refer to our nature as human beings. However, as we shall see, there are good reasons to doubt the truth of animalism, and those reasons also suggest that the two answers to ‘What is a person?’ are largely independent.

The methodology of thought-experiments

As argued above, the best way to answer the nature question is by answering the identity question. But how should we answer the iden-

tity question? Evidently, consideration of ordinary cases will not help us to decide the issue. For example, when I judge that the lecturer before me now is identical to the lecturer who began speaking an hour ago, I typically make this judgement of identity under pretty much optimal conditions. In such a case, I can observe that the earlier person is both physically and psychologically continuous with the later person. The very same brain and body has persisted for one hour, and that brain (we may suppose) has directly supported the very same beliefs, character, desires and memories (with only very slight changes).

In this everyday case, my judgement of identity is based on the obtaining of both physical and psychological continuities. Reflection on such a case evidently will not help to determine which continuity (if either) is more important or central to the identity of a person over time. We will need to consider *thought-experiments* in which these continuities come apart. The events depicted in the thought-experiments in this book are all technically impossible at present, and may always be so. But we have no reason to think that any of the thought-experiments is physically impossible (that is, inconsistent with the laws of nature). And, certainly, none is logically impossible.

The use of thought-experiments in philosophy has been subject to a number of criticisms. It has been claimed that we should not take our intuitions about thought-experiments as guides to philosophical truth, since such intuitions may be prejudiced and unreliable. This criticism is, I think, over-stated. For one thing, it ignores the frequent and legitimate use of thought-experiments in virtually all traditional areas of philosophy (most notably, for example, in theories of knowledge and in ethics).

Second, and more important, thought-experiments can be useful in understanding the structure of a concept and the relative importance of its different strands, provided that there is general agreement about the best description of the thought-experiment. It's true that some philosophers have tried to gain mileage from thought-experiments in the absence of such general agreement. But it would be unwarranted to infer from the existence of such abuses that thought-experiments can never perform any useful function in philosophy.

Thus, consider Wittgenstein's verdict on the following thought-experiment:

Imagine a man whose memories on the even days of his life comprise the events of all these days, skipping entirely what

happened on the odd days. On the other hand, he remembers on an odd day what happened on previous odd days, but his memory then skips the even days without a feeling of discontinuity.... Are we bound to say that here two persons are inhabiting the same body? That is, is it right to say that there are, and wrong to say that there aren't, or vice versa? Neither. For the *ordinary* use of the word 'person' is what one might call a composite use suitable under ordinary circumstances. If I assume, as I do, that these circumstances are changed, the application of the term 'person' or 'personality' has thereby changed; and if I wish to preserve this term and give it a use analogous to its former use, I am at liberty to choose between many uses, that is, between many different kinds of analogy. One might say in such a case that the term 'personality' hasn't got one legitimate heir only.¹³

Wittgenstein has here described a nice case where neither the answer 'Only one person occupies the body throughout' nor the answer 'Two people alternately occupy the body' are correct or satisfactory. That is, Wittgenstein's thought-experiment exploits the vagueness or indeterminacy of our concepts *person* and *same person*. We may choose to stipulate a more precise meaning for the term 'person', allowing us to say, for example, 'the case involves two people'. But, if we do so, we must be aware that that is what we are doing. We are not reading-off a definite answer from our concept of a person—a concept clearly not designed to yield a yes-or-no answer to questions of personal identity in all possible cases. (The concept *person* is vague in another way too: it can sometimes be vague whether a given entity (for example, a neonate) is a person. But such vagueness is not relevant to the present discussion.)

However, none of this tells against the methodology of thought-experiments. It just shows that, in some thought-experiments, there is no definite answer to questions of personal identity. This is a result that no one ought to dispute.

In this book, we will appeal to a number of thought-experiments to help decide the identity question. The point of these thought-experiments is to enable us to extract a core (that is, minimally controversial) set of common-sense beliefs about the conditions of personal identity over time. In all these thought-experiments, unlike in the Cartesian thought-experiment of a soul floating free of a dead body, we respect the empirically supported fact that states of the mind depend

upon states of the brain. This gives our thought-experiments a grounding that Descartes' conceivings lacked.

Here, briefly, are some of the thought-experiments which will feature in subsequent chapters:

Brain Transplant

My brain is removed from my body, kept alive, and then hooked up inside a new skull and body, exactly similar to my old skull and body. My old body is destroyed. The resulting person has my brain and a new body. Since my brain directly supports my mental life, the new person is psychologically continuous with me.¹⁴

Scattered Existence

My brain is removed from my body and stored in a vat. It is 'connected' to my now brainless body by radio links. I can 'see' and 'hear' appropriately placed objects in the vicinity of my body, yet my brain is hundreds of miles from my body. Suddenly, an avalanche destroys my body. I am still conscious, but receiving no sensory input....¹⁵

Bionic Replacement

My brain develops cancer. Technology has reached the stage where any human brain function can be mimicked by an appropriate collection of silicon chips. So my surgeons offer to carry out the following operation: they will gradually replace all my biological brain with silicon parts. I will end up with an entirely bionic brain. The new bionic brain will subserve the very same psychological functions as the original. In other words, I will be psychologically continuous with the resulting individual composed of a flesh and blood body and a bionic brain.

Teletransportation

On Earth, I step into the scanner. The function of the scanner is to create an exact atom-for-atom blueprint of me, and then painlessly to destroy me by vaporisation. On the surface of a distant planet, out of different matter, a replicator receives the blueprint and creates an exact replica of me. The replica looks like me, and has all my physical characteristics. He also has all my mental characteristics,

since mental properties depend significantly on physical properties of the brain, and the replicated brain is physically identical to my original brain. Yet my replica has no material substance in common with me.

Branch-line

I am replicated on the distant planet's surface, but the scanner on Earth is now programmed not to vaporise me. However, the operation of the scanner causes me to have cardiac failure on Earth. I am still conscious, and know that I have only a few days to live.¹⁶

Accident

I am in a horrendous car accident, and suffer massive brain damage. In fact, my psychological life has been completely destroyed, but my body and brain are artificially kept alive. The surgeons find a way to make my brain function again. But complete re-training is necessary. It takes years to advance from the psychological level of a newborn infant to that of a normal adult. The resulting person is quite unlike me psychologically. He and I are not at all psychologically continuous.¹⁷

Indeterminacy

An alteration machine changes me physically and psychologically. My brain is refigured so that roughly half of my memories, beliefs, desires, and character traits are replaced with new and very different ones. It is vague or indeterminate whether I am psychologically continuous with the resulting person.

Fission

My body is riddled with cancer. The surgeons want to try out a new technique: hemisphere transplant. They have two brainless donor bodies available, cloned years ago from my body. Each of my two brain hemispheres is removed and placed in its own body. Two persons result. Since I am one of the few people whose brain hemispheres are functionally equivalent (that is, they support the very same mental capacities), both resulting persons will think they are me, and they will both have my character, apparent memories, and all my other psychological features.¹⁸

THE PROBLEM AND ITS PLACE IN PHILOSOPHY

In these thought-experiments, the first question to ask is: What has happened to me? Have I survived? Have I died? Or is there no definite answer? We shall address these questions in coming chapters.