

At last the horizon appears free to us again, even granted that it is not bright; at last our ships may venture out again, venture out to face any danger; all the daring of the lover of knowledge is permitted again; the sea, *our sea*, lies open again; perhaps there has never yet been such an 'open sea'.

NIETZSCHE, p. 448

REASONS AND PERSONS

BY

DEREK PARFIT

CLARENDON PRESS · OXFORD

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi Kuala Lumpur
Madrid Melbourne Mexico City Nairobi New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Derek Parfit 1984

The moral rights of the author have been asserted

Database right Oxford University Press (maker)

First published 1984

First issued in paperback (with corrections) 1986

Reprinted with further corrections 1987

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose this same condition on any acquirer

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Parfit, Derek.

Reasons and persons.

Includes bibliographical references and index.

1. Ethics 2. Rationalism. 3. Self. I. Title.

BJ1012.P39 1984 170 83-15139

ISBN 978-0-19-824908-5 (pbk)

To my parents

Drs. Jessie and Norman Parfit

and my sisters

Theodora and Joanna

Printed in Great Britain

on acid-free paper by

CPI Antony Rowe, Chippenham, Wiltshire

CHAPTER 4 · THEORIES THAT ARE DIRECTLY
SELF-DEFEATING

32	In Prisoner's Dilemmas, Does S Fail in Its Own Terms?	87
33	Another Weak Defence of Morality	88
34	Intertemporal Dilemmas	91
35	A Weak Defence of S	92
36	How Common-Sense Morality Is Directly Self-Defeating	93
37	The Five Parts of a Moral Theory	95
38	How We Can Revise Common-Sense Morality so that It Would Not Be Self-Defeating	98
39	Why We Ought to Revise Common-Sense Morality	100
40	A Simpler Revision	103

CHAPTER 5 · CONCLUSIONS

41	Reducing the Distance between M and C	111
42	Towards a Unified Theory	111
43	Work to be Done	112
44	Another Possibility	113

PART TWO · RATIONALITY AND TIME

CHAPTER 6 · THE BEST OBJECTION TO THE
SELF-INTEREST THEORY

45	The Present-aim Theory	117
46	Can Desires Be Intrinsically Irrational, or Rationally Required?	117
47	Three Competing Theories	120
48	Psychological Egoism	126
49	The Self-interest Theory and Morality	127
50	My First Argument	129
51	The S-Theorist's First Reply	130
52	Why Temporal Neutrality Is Not the Issue Between S and P	132

CHAPTER 7 · THE APPEAL TO FULL RELATIVITY

53	The S-Theorist's Second Reply	133
54	Sidgwick's Suggestions	137
55	How S Is Incompletely Relative	137
56	How Sidgwick Went Astray	140
57	The Appeal Applied at a Formal Level	141
58	The Appeal Applied to Other Claims	142

CHAPTER 8 · DIFFERENT ATTITUDES TO TIME

59	Is It Irrational to Give No Weight to One's Past Desires?	144
60	Desires that Depend on Value Judgements or Ideals	149
61	Mere Past Desires	153
62	Is It Irrational To Care Less About One's Further Future?	156
63	A Suicidal Argument	158
64	Past or Future Suffering	163
65	The Direction of Causation	165
66	Temporal Neutrality	168
67	Why We Should Not Be Biased towards the Future	170

68	Time's Passage	177
69	An Asymmetry	181
70	Conclusions	184

CHAPTER 9 · WHY WE SHOULD REJECT S

71	The Appeal to Later Regrets	187
72	Why a Defeat for Proximus is Not a Victory for S	187
73	The Appeal to Inconsistency	188
74	Conclusions	189

PART THREE · PERSONAL IDENTITY

CHAPTER 10 · WHAT WE BELIEVE OURSELVES TO BE

75	Simple Teletransportation and the Branch-Line Case	199
76	Qualitative and Numerical Identity	200
77	The Physical Criterion of Personal Identity	201
78	The Psychological Criterion	202
79	The Other Views	204

CHAPTER 11 · HOW WE ARE NOT WHAT WE BELIEVE

80	Does Psychological Continuity Presuppose Personal Identity?	219
81	The Subject of Experiences	219
82	How a Non-Reductionist View Might Have Been True	223
83	Williams's Argument against the Psychological Criterion	227
84	The Psychological Spectrum	229
85	The Physical Spectrum	231
86	The Combined Spectrum	234

CHAPTER 12 · WHY OUR IDENTITY IS NOT WHAT MATTERS

87	Divided Minds	245
88	What Explains the Unity of Consciousness?	245
89	What Happens When I Divide?	248
90	What Matters When I Divide?	253
91	Why There Is No Criterion of Identity that Can Meet Two Plausible Requirements	261
92	Wittgenstein and Buddha	266
93	Am I Essentially My Brain?	273
94	Is the True View Believable?	273

CHAPTER 13 · WHAT DOES MATTER

95	Liberation From the Self	281
96	The Continuity of the Body	281
97	The Branch-Line Case	282
98	Series-Persons	287
99	Am I a Token or a Type?	289
100	Partial Survival	293
101	Successive Selves	298

CHAPTER 14 · PERSONAL IDENTITY AND RATIONALITY

102	The Extreme Claim	307
103	A Better Argument against the Self-interest Theory	307

PART THREE

PERSONAL IDENTITY

WHAT WE BELIEVE OURSELVES TO BE

I enter the Teletransporter. I have been to Mars before, but only by the old method, a space-ship journey taking several weeks. This machine will send me at the speed of light. I merely have to press the green button. Like others, I am nervous. Will it work? I remind myself what I have been told to expect. When I press the button, I shall lose consciousness, and then wake up at what seems a moment later. In fact I shall have been unconscious for about an hour. The Scanner here on Earth will destroy my brain and body, while recording the exact states of all of my cells. It will then transmit this information by radio. Travelling at the speed of light, the message will take three minutes to reach the Replicator on Mars. This will then create, out of new matter, a brain and body exactly like mine. It will be in this body that I shall wake up.

Though I believe that this is what will happen, I still hesitate. But then I remember seeing my wife grin when, at breakfast today, I revealed my nervousness. As she reminded me, she has been often teletransported, and there is nothing wrong with *her*. I press the button. As predicted, I lose and seem at once to regain consciousness, but in a different cubicle. Examining my new body, I find no change at all. Even the cut on my upper lip, from this morning's shave, is still there.

Several years pass, during which I am often Teletransported. I am now back in the cubicle, ready for another trip to Mars. But this time, when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I leave the cubicle, and say to the attendant: 'It's not working. What did I do wrong?'

'It's working', he replies, handing me a printed card. This reads: 'The New Scanner records your blueprint without destroying your brain and body. We hope that you will welcome the opportunities which this technical advance offers.'

The attendant tells me that I am one of the first people to use the New Scanner. He adds that, if I stay for an hour, I can use the Intercom to see and talk to myself on Mars.

'Wait a minute', I reply, 'If I'm here I can't *also* be on Mars'.

Someone politely coughs, a white-coated man who asks to speak to

me in private. We go to his office, where he tells me to sit down, and pauses. Then he says: 'I'm afraid that we're having problems with the New Scanner. It records your blueprint just as accurately, as you will see when you talk to yourself on Mars. But it seems to be damaging the cardiac systems which it scans. Judging from the results so far, though you will be quite healthy on Mars, here on Earth you must expect cardiac failure within the next few days.'

The attendant later calls me to the Intercom. On the screen I see myself just as I do in the mirror every morning. But there are two differences. On the screen I am not left-right reversed. And, while I stand here speechless, I can see and hear myself, in the studio on Mars, starting to speak.

What can we learn from this imaginary story? Some believe that we can learn little. This would have been Wittgenstein's view.¹ And Quine writes: 'The method of science fiction has its uses in philosophy, but. . . I wonder whether the limits of the method are properly heeded. 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.'²

This criticism 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 our 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. We discover our beliefs about the nature of personal identity over time. Though our beliefs are revealed most clearly when we consider imaginary cases, these beliefs also cover actual cases, and our own lives. In Part Three of this book I shall argue that some of these beliefs are false, then suggest how and why this matters.

75. SIMPLE TELETRANSPORTATION AND THE BRANCH-LINE CASE

At the beginning of my story, the Scanner destroys my brain and body. My blueprint is beamed to Mars, where another machine makes an organic *Replica* of me. My Replica thinks that he is me, and he seems to remember living my life up to the moment when I pressed the green button. In every other way, both physically and psychologically, we are exactly similar. If he returned to Earth, everyone would think that he was me.

Simple Teletransportation, as just described, is a common feature in science fiction. And it is believed, by some readers of this fiction, merely to be the fastest way of travelling. They believe that my Replica *would be me*. Other science fiction readers, and some of the characters in this fiction, take a different view. They believe that, when I press the green button, I die. My Replica is *someone else*, who has been made to be exactly like me.

This second view seems to be supported by the end of my story. The New Scanner does not destroy my brain and body. Besides gathering the information, it merely damages my heart. While I am in the cubicle, with the green button pressed, nothing seems to happen. I walk out, and learn that in a few days I shall die. I later talk, by two-way television, to my Replica on Mars. Let us continue the story. Since my Replica knows that I am about to die, he tries to console me with the same thoughts with which I recently tried to console a dying friend. It is sad to learn, on the receiving end, how unconsoling these thoughts are. My Replica then assures me that he will take up my life where I leave off. He loves my wife, and together they will care for my children. And he will finish the book that I am writing. Besides having all of my drafts, he has all of my intentions. I must admit that he can finish my book as well as I could. All these facts console me a little. Dying when I know that I shall have a Replica is not quite as bad as, simply, dying. Even so, I shall soon lose consciousness, forever.

In Simple Teletransportation, I am destroyed before I am Replicated. This makes it easier to believe that this *is* a way of travelling—that my Replica *is* me. At the end of my story, my life and that of my Replica overlap. Call this the *Branch-Line Case*. In this case, I cannot hope to travel on the *Main Line*, waking up on Mars with forty years of life ahead. I shall stay on the Branch-Line, here on Earth, which ends a few days later. Since I can talk to my Replica, it seems clear that he is *not* me. Though he is exactly like me, he is one person, and I am another. When I pinch myself, he feels nothing. When I have my heart attack, he will again feel nothing. And when I am dead he will live for another forty years.

If we believe that my Replica is not me, it is natural to assume that my prospect, on the Branch Line, is almost as bad as ordinary death. I shall deny this assumption. As I shall argue later, being destroyed and Replicated is about as good as ordinary survival. I can best defend this claim, and the wider view of which it is part, after discussing the past debate about personal identity.

76. QUALITATIVE AND NUMERICAL IDENTITY

There are two kinds of sameness, or identity. I and my Replica are *qualitatively identical*, or exactly alike. But we may not be *numerically identical*, or one and the same person. Similarly, two white billiard balls are not numerically but may be qualitatively identical. If I paint one of these balls red, it will cease to be qualitatively identical with itself as it was. But the red ball that I later see and the white ball that I painted red are numerically identical. They are one and the same ball.

We might say, of someone, 'After his accident, he is no longer the same person'. This is a claim about both kinds of identity. We claim that *he*, the same person, is *not* now the same person. This is not a contradiction. We

merely mean that this person's character has changed. This numerically identical person is now qualitatively different.

When we are concerned about our future, it is our numerical identity that we are concerned about. I may believe that, after my marriage, I shall not be the same person. But this does not make marriage death. However much I change, I shall still be alive if there will be some person living who will *be* me.

Though our chief concern is our numerical identity, psychological changes matter. Indeed, on one view, certain kinds of qualitative change destroy numerical identity. If certain things happen to me, the truth might not be that I become a very different person. The truth might be that I cease to exist—that the resulting person is someone else.

77. THE PHYSICAL CRITERION OF PERSONAL IDENTITY

There has been much debate about the nature both of persons and of personal identity over time. It will help to distinguish these questions:

- (1) What is the nature of a person?
- (2) What makes a person at two different times one and the same person? What is necessarily involved in the continued existence of each person over time?

The answer to (2) can take this form: '*X* today is one and the same person as *Y* at some past time *if and only if* . . .' Such an answer states the *necessary and sufficient conditions* for personal identity over time.

In answering (2) we shall also partly answer (1). The necessary features of our continued existence depend upon our nature. And the simplest answer to (1) is that, to be a person, a being must be self-conscious, aware of its identity and its continued existence over time.

We can also ask

- (3) What is in fact involved in the continued existence of each person over time?

Since our continued existence has features that are not necessary, the answer to (2) is only part of the answer to (3). For example, having the same heart and the same character are not necessary to our continued existence, but they are usually part of what this existence involves.

Many writers use the ambiguous phrase 'the criterion of identity over time'. Some mean by this 'our way of telling whether some present object is identical with some past object'. But I shall mean *what this identity necessarily involves, or consists in*.

In the case of most physical objects, on what I call the *standard view*, the criterion of identity over time is the spatio-temporal physical continuity of this object. This is something that we all understand, even if we fail to understand the description I shall now give. In the simplest case of physical continuity, like that of the Pyramids, an apparently static object continues to exist. In another simple case, like that of the Moon, an object moves in a regular way. Many objects move in less regular ways, but they still trace physically continuous spatio-temporal paths. Suppose that the billiard ball that I painted red is the same as the white ball with which last year I made a winning shot. On the standard view, this is true only if this ball traced such a continuous path. It must be true (1) that there is a line through space and time, starting where the white ball rested before I made my winning shot, and ending where the red ball now is, (2) that at every point on this line there was a billiard ball, and (3) that the existence of a ball at each point on this line was in part caused by the existence of a ball at the immediately preceding point.³

Some kinds of thing continue to exist even though their physical continuity involves great changes. A Camberwell Beauty is first an egg, then a caterpillar, then a chrysalis, then a butterfly. These are four stages in the physically continuous existence of a single organism. Other kinds of thing cannot survive such great changes. Suppose that an artist paints a self-portrait and then, by repainting, turns this into a portrait of his father. Even though these portraits are more similar than a caterpillar and a butterfly, they are not stages in the continued existence of a single painting. The self-portrait is a painting that the artist destroyed. In a general discussion of identity, we would need to explain why the requirement of physical continuity differs in such ways for different kinds of thing. But we can ignore this here.

Can there be gaps in the continued existence of a physical object? Suppose that I have the same gold watch that I was given as a boy even though, for a month, it lay disassembled on a watch-repairer's shelf. On one view, in the spatio-temporal path traced by this watch there was not at every point a watch, so my watch does not have a history of full physical continuity. But during the month when my watch was disassembled, and did not exist, all of its parts had histories of full continuity. On another view, even when it was disassembled, my watch existed.

Another complication again concerns the relation between a complex thing and the various parts of which it is composed. It is true of some of these things, though not true of all, that their continued existence need not involve the continued existence of their components. Suppose that a wooden ship is repaired from time to time while it is floating in harbour, and that after fifty years it contains none of the bits of wood out of which it was first built. It is still one and the same ship, because, as a ship, it has

displayed throughout these fifty years full physical continuity. This is so despite the fact that it is now composed of quite different bits of wood. These bits of wood might be qualitatively identical to the original bits, but they are not one and the same bits. Something similar is partly true of a human body. With the exception of some brain cells, the cells in our bodies are replaced with new cells several times in our lives.

I have now described the physical continuity which, on the standard view, makes a physical object one and the same after many days or years. This enables me to state one of the rival views about personal identity. On this view, what makes me the same person over time is that I have the same brain and body. The criterion of my identity over time—or what this identity involves—is the physical continuity, over time, of my brain and body. I shall continue to exist if and only if this particular brain and body continue both to exist and to be the brain and body of a living person.

This is the simplest version of this view. There is a better version. This is

The Physical Criterion: (1) What is necessary is not the continued existence of the whole body, but the continued existence of *enough* of the brain to be the brain of a living person. *X* today is one and the same person as *Y* at some past time if and only if (2) enough of *Y*'s brain continued to exist, and is now *X*'s brain, and (3) this physical continuity has not taken a 'branching' form. (4) Personal identity over time just consists in the holding of facts like (2) and (3).

(1) is clearly true in certain actual cases. Some people continue to exist even though they lose, or lose the use of, much of their bodies. (3) will be explained later.

Those who believe in the Physical Criterion would reject Teletransportation. They would believe this to be a way, not of travelling, but of dying. They would also reject, as inconceivable, reincarnation. They believe that someone cannot have a life after death, unless he lives this life in a resurrection of the very same, physically continuous body. This is why some Christians insist that they be buried. They believe that if, like Greek and Trojan heroes, they were burnt on funeral pyres, and their ashes scattered, not even God could bring them to life again. God could create only a Replica, someone else who was exactly like them. Other Christians believe that God could resurrect *them* if He reassembled their bodies out of the bits of matter that, when they were last alive, made up their bodies. This would be like the reassembly of my gold watch.⁴

78. THE PSYCHOLOGICAL CRITERION

Some people believe in a kind of psychological continuity that resembles physical continuity. This involves the continued existence of a purely mental

entity, or thing—a soul, or spiritual substance. I shall return to this view. But I shall first explain another kind of psychological continuity. This is less like physical continuity, since it does not consist in the continued existence of some entity. But this other kind of psychological continuity involves only facts with which we are familiar.

What has been most discussed is the continuity of memory. This is because it is memory that makes most of us aware of our own continued existence over time. The exceptions are the people who are suffering from amnesia. Most amnesiacs lose only two sets of memories. They lose all of their memories of having particular past experiences—or, for short, their *experience memories*. They also lose some of their memories about facts, those that are about their own past lives. But they remember other facts, and they remember how to do different things, such as how to speak, or swim.

Locke suggested that experience-memory provides the criterion of personal identity.⁵ Though this is not, on its own, a plausible view, I believe that it can be part of such a view. I shall therefore try to answer some of Locke's critics.

Locke claimed that someone cannot have committed some crime unless he now remembers doing so. We can understand a reluctance to punish people for crimes that they cannot remember. But, taken as a view about what is involved in a person's continued existence, Locke's claim is clearly false. If it was true, it would not be possible for someone to forget any of the things that he once did, or any of the experiences that he once had. But this is possible. I cannot now remember putting on my shirt this morning.

There are several ways to extend the experience-memory criterion so as to cover such cases. I shall appeal to the concept of an overlapping chain of experience-memories. Let us say that, between *X* today and *Y* twenty years ago, there are *direct memory connections* if *X* can now remember having some of the experiences that *Y* had twenty years ago. On Locke's view, only this makes *X* and *Y* one and the same person. But even if there are *no* such direct memory connections, there may be *continuity of memory* between *X* now and *Y* twenty years ago. This would be so if between *X* now and *Y* at that time there has been an overlapping chain of direct memories. In the case of most adults, there would be such a chain. In each day within the last twenty years, most of these people remembered some of their experiences on the previous day. On the revised version of Locke's view, some present person *X* is the same as some past person *Y* if there is between them continuity of memory.

This revision meets one objection to Locke's view. We should also revise the view so that it appeals to other facts. Besides direct memories, there are several other kinds of direct psychological connection. One such connection is that which holds between an intention and the later act in which this intention is carried out. Other such direct connections are those which hold when a belief, or a desire, or any other psychological feature, continues to be had.

I can now define two general relations:

Psychological connectedness is the holding of particular direct psychological connections.

Psychological continuity is the holding of overlapping chains of *strong* connectedness.

Of these two general relations, connectedness is more important both in theory and in practice. Connectedness can hold to any degree. Between X today and Y yesterday there might be several thousand direct psychological connections, or only a single connection. If there was only a single connection, X and Y would not be, on the revised Lockean View, the same person. For X and Y to be the same person, there must be over every day *enough* direct psychological connections. Since connectedness is a matter of degree, we cannot plausibly define precisely what counts as enough. But we can claim that there is enough connectedness if the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person.⁶ When there are enough direct connections, there is what I call *strong* connectedness.

Could this relation be the criterion of personal identity? A relation *F* is *transitive* if it is true that, if *X* is *F*-related to *Y*, and *Y* is *F*-related to *Z*, *X* and *Z* *must* be *F*-related. Personal identity is a transitive relation. If Bertie was one and the same person as the philosopher Russell, and Russell was one and the same person as the author of *Why I Am Not a Christian*, this author and Bertie must be one and the same person.

Strong connectedness is *not* a transitive relation. I am now strongly connected to myself yesterday, when I was strongly connected to myself two days ago, when I was strongly connected to myself three days ago, and so on. It does not follow that I am now strongly connected to myself twenty years ago. And this is not true. Between me now and myself twenty years ago there are many fewer than the number of direct psychological connections that hold over any day in the lives of nearly all adults. For example, while most adults have many memories of experiences that they had in the previous day, I have few memories of experiences that I had on any day twenty years ago.

By 'the criterion of personal identity over time' I mean what this identity *necessarily involves or consists in*. Because identity is a transitive relation, the criterion of identity must also be a transitive relation. Since strong connectedness is not transitive, it cannot be the criterion of identity. And I have just described a case in which this is clear. I am the same person as myself twenty years ago, though I am not now strongly connected to myself then.

Though a defender of Locke's view cannot appeal to psychological connectedness, he can appeal to psychological continuity, which is transitive. He can appeal to

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

As with the Physical Criterion, (4) will be explained later.

There are three versions of the Psychological Criterion. These differ over the question of what is the *right* kind of cause. On the *Narrow* version, this must be the *normal* cause. On the *Wide* version, this could be *any reliable* cause. On the *Widest* version, the cause could be *any* cause.

The Narrow Psychological Criterion uses words in their ordinary sense. Thus I remember having an experience only if

(1) I seem to remember having an experience,

(2) I did have this experience,

and

(3) my apparent memory is causally dependent, in the normal way, on this past experience.

That we need condition (3) can be suggested with an example. Suppose that I am knocked unconscious in a climbing accident. After I recover, my fellow-climber tells me what he shouted just before I fell. In some later year, when my memories are less clear, I might seem to remember the experience of hearing my companion shout just before I fell. And it might be true that I did have just such an experience. But though conditions (1) and (2) are met, we should not believe that I am remembering that past experience. It is a well-established fact that people can never remember their last few experiences before they were knocked unconscious. We should therefore claim that my apparent memory of hearing my companion shout is not a real memory of that past experience. This apparent memory is not causally dependent in the right way on that past experience. I have this apparent memory only because my companion later told me what he shouted.

Similar remarks apply to the other kinds of continuity, such as continuity of character. On the Narrow Psychological Criterion, even if someone's character radically changes, there is continuity of character if these changes have one of several normal causes. Some changes of character are deliberately brought about; others are the natural consequence of growing older; others are the natural response to certain kinds of experience. But there would not be continuity of character if radical and unwanted changes were produced by abnormal interference, such as direct tampering with the brain.

Though it is memory that makes us aware of our own continued existence over time, the various other continuities have great importance. We may believe that they have enough importance to provide personal identity even in the absence of memory. We shall then claim, what Locke denied, that a person continues to exist even if he suffers from complete amnesia.

Besides the Narrow version, I described the two Wide versions of the Psychological Criterion. These versions extend the senses of several words. On the ordinary sense of 'memory', a memory must have its normal cause. The two Wide Psychological Criteria appeal to a wider sense of 'memory', which allows either any reliable cause, or any cause. Similar claims apply to the other kinds of direct psychological connection. To simplify my discussion of these three Criteria, I shall use 'psychological continuity' in its widest sense, that allows this continuity to have *any* cause.

If we appeal to the Narrow Version, which insists on the normal cause, the Psychological Criterion coincides in most cases with the Physical Criterion. The normal causes of memory involve the continued existence of the brain. And some or all of our psychological features depend upon states or events in our brains. The continued existence of a person's brain is at least part of the normal cause of psychological continuity. On the Physical Criterion, a person continues to exist if and only if (a) there continues to exist *enough* of this person's brain so that it remains the brain of a living person, and (b) there has been no branching in this physical continuity. (a) and (b) are claimed to be the necessary and sufficient conditions for this person's identity, or continued existence, over time. On the Narrow Psychological Criterion, (a) is necessary, but not sufficient. A person continues to exist if and only if (c) there is psychological continuity, (d) this continuity has its normal cause, and (e) it has not taken a branching form. (a) is required as part of the normal cause of psychological continuity.

Reconsider the start of my imagined story, where my brain and body are destroyed. The Scanner and the Replicator produce a person who has a new but exactly similar brain and body, and who is psychologically continuous with me as I was when I pressed the green button. The cause of this continuity is, though unusual, reliable. On both the Physical Criterion and the Narrow Psychological Criterion, my Replica would *not* be me. On the two Wide Criteria, he *would* be me.

I shall argue that we need not decide between these three versions of the Psychological Criterion. A partial analogy may suggest why. Some people go blind because of damage to their eyes. Scientists are now developing artificial eyes. These involve a glass or plastic lens, and a micro-computer which sends through the optic nerve electrical patterns like those that are sent through this nerve by a natural eye. When such artificial eyes are more advanced, they might give to someone who has gone blind visual experiences just like those that he used to have. What he seems to see would correspond to what is in fact

before him. And his visual experiences would be causally dependent, in this new but reliable way, on the light-waves coming from the objects that are before him.

Would this person be *seeing* these objects? If we insist that seeing must involve the normal cause, we would answer No. But even if this person cannot see, what he has is *just as good as* seeing, both as a way of knowing what is within sight, and as a source of visual pleasure. If we accept the Psychological Criterion, we could make a similar claim. If psychological continuity does not have its normal cause, it may not provide personal identity. But we can claim that, even if this is so, what it provides is *as good as* personal identity.

79. THE OTHER VIEWS

I am asking what is the criterion of personal identity over time—what this identity involves, or consists in. I first described the spatio-temporal physical continuity that, on the standard view, is the criterion of identity of physical objects. I then described two views about personal identity, the Physical and Psychological Criteria.

There is a natural but false assumption about these views. Many people believe in what is called *Materialism*, or *Physicalism*. This is the view that that there are no purely mental objects, states, or events. On one version of Physicalism, every mental event is just a physical event in some particular brain and nervous system. There are other versions. Those who are not Physicalists are either *Dualists* or *Idealists*. Dualists believe that mental events are *not* physical events. This can be so even if all mental events are causally dependent on physical events in a brain. Idealists believe that all states and events are, when understood correctly, purely mental. Given these distinctions, we may assume that Physicalists must accept the Physical Criterion of personal identity.

This is not so. Physicalists could accept the Psychological Criterion. And they could accept the version that allows any reliable cause, or any cause. They could thus believe that, in Simple Teletransportation, my Replica would be me. They would here be rejecting the Physical Criterion.⁸

These criteria are not the only views about personal identity. I shall now describe some of the other views that are either sufficiently plausible, or have enough supporters, to be worth considering. This description may be hard to follow; but it will give a rough idea of what lies ahead. If much of this summary seems either obscure or trivial, do not worry.

I start with a new distinction. On the Physical Criterion, personal identity over time just involves the physically continuous existence of enough of a

brain so that it remains the brain of a living person. On the Psychological Criterion, personal identity over time just involves the various kinds of psychological continuity, with the right kind of cause. These views are both *Reductionist*. They are Reductionist because they claim

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

They may also claim

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

It may seem that (2) could not be true. When we describe the psychological continuity that unifies some person's mental life, we must mention this person, and many other people, in describing the *content* of many thoughts, desires, intentions, and other mental states. But mentioning this person in this way does not involve either asserting that these mental states are had by this person, or asserting that this person exists. These claims need further arguments, which I shall later give.

Our view is *Non-Reductionist* if we reject both of the two Reductionist claims.

Many Non-Reductionists believe that *we are separately existing entities*. On this view, personal identity over time does not just consist in physical and/or psychological continuity. It involves a further fact. A person is a separately existing entity, distinct from his brain and body, and his experiences. On the best-known version of this view, a person is a *purely mental* entity: a Cartesian Pure Ego, or spiritual substance. But we might believe that a person is a separately existing *physical* entity, of a kind that is not yet recognised in the theories of contemporary physics.

There is another Non-Reductionist View. This view denies that we are separately existing entities, distinct from our brains and bodies, and our experiences. But this view claims that, though we are not separately existing entities, personal identity *is* a further fact, which does not just consist in physical and/or psychological continuity. I call this the *Further Fact View*.

I shall now draw some more distinctions. The Physical and Psychological Criteria are versions of the Reductionist View; and there are different versions of each criterion. But what is necessarily involved in a person's continued existence is less than what is in fact involved. So while believers in

the different criteria disagree about imaginary cases, they agree about what is in fact involved in the continued existence of most actual people. They would start to disagree only if, for example, people began to be Teletransported.

On the Reductionist View, each person's existence just involves the existence of a brain and body, the doing of certain deeds, the thinking of certain thoughts, the occurrence of certain experiences, and so on. It will help to extend the ordinary sense of the word 'event'. I shall use this word to cover even such *boring* events as the continued existence of a belief, or a desire. This use makes the Reductionist View simpler to describe. And it avoids what I believe to be one misleading implication of the words 'mental state'. While a state must be a state *of* some entity, this is not true of an event. Given this extended use of the word 'event', all Reductionists would accept

- (3) A person's existence just consists in the existence of a brain and body, and the occurrence of a series of interrelated physical and mental events.

Some Reductionists claim

- (4) A person *just is* a particular brain and body, and such a series of interrelated events.

Other Reductionists claim

- (5) A person is an entity that is *distinct* from a brain and body, and such a series of events.

On this version of the Reductionist View, a person is not merely a composite object, with these various components. A person is an entity that *has* a brain and body, and *has* particular thoughts, desires, and so on. But, though (5) is true, a person is not a *separately existing* entity. Though (5) is true, (3) is also true.

This version of Reductionism may seem self-contradictory. (3) and (5) may seem to be inconsistent. It may help to consider Hume's analogy: 'I cannot compare the soul more properly to anything than to a republic, or commonwealth.'⁹ Most of us are Reductionists about nations. We would accept the following claims: Nations exist. Ruritania does not exist, but France does. Though nations exist, a nation is not an entity that exists separately, apart from its citizens and its territory. We would accept

- (6) A nation's existence just involves the existence of its citizens,

living together in certain ways, on its territory.

Some claim

- (7) A nation just *is* these citizens and this territory.

Others claim

- (8) A nation is an entity that is distinct from its citizens and its territory.

For reasons that I give in Appendix D, we may believe that (6) and (8) are not inconsistent. If we believe this, we may accept that there is no inconsistency between the corresponding claims (3) and (5). We may thus agree that the version of Reductionism expressed in (3) and (5) is a consistent view. If this version is consistent, as I believe, it is the better version. It stays closer to our actual concept of a person. But in most of what follows we can ignore the difference between these two versions.

Besides claiming (1) and (2), Reductionists might also claim

- (9) Though persons exist, we could give a *complete* description of reality *without* claiming that persons exist.

I call this the view *that a complete description could be impersonal*.

This view may also seem to be self-contradictory. If persons exist, and a description of what exists fails to mention persons, how can this description be complete?

A Reductionist could give the following reply. Suppose that an object has two names. This is true of the planet that is called both *Venus* and the *Evening Star*. In our description of what exists, we could claim that Venus exists. Our description could then be complete even though we do not claim that the Evening Star exists. We need not make this claim because, using its other name, we have already claimed that this object exists.

A similar claim applies when some fact can be described in two ways. Some Reductionists accept (4), the claim that a person just is a particular brain and body, and a series of interrelated physical and mental events. If this is what a person *is*, we can describe this fact by claiming either

- (10) that there exists a particular brain and body, and a particular series of interrelated physical and mental events.

or

- (11) that a particular person exists,

If (10) and (11) are two ways of describing the *same* fact, a complete description need not make *both* claims. It would be enough to make claim (10). Though this person exists, a complete description need not claim that he

exists, since this fact has already been reported in claim (10).

Other Reductionists accept (5), the claim that a person is distinct from his brain and body, and his acts, thoughts, and other physical and mental events. On this version of Reductionism, claim (10) does not describe the very same fact that claim (11) describes. But claim (10) may *imply* claim (11). More cautiously, given our understanding of the concept of a person, if we know that (10) is true, we shall know that (11) is true. These Reductionists can say that, if our description of reality either states or implies, or enables us to know about, the existence of everything that exists, our description is *complete*. This claim is not as clearly true as the claim that a complete description need not give two descriptions of the same fact. But this claim seems plausible. If it is justified, and the Reductionist View is true, these Reductionists can completely describe reality without claiming that persons exist.¹¹

My claims about Reductionism draw distinctions that, in this abstract form, are hard to grasp. But there are other ways of discovering whether we are Reductionists in our view about some kind of thing. If we accept a Reductionist View, we shall believe that the identity of such a thing may be, in a quite unpuzzling way, *indeterminate*. If we do *not* believe this, we are probably Non-Reductionists about this kind of thing.

Consider, for example, clubs. Suppose that a certain club exists for several years, holding regular meetings. The meetings then cease. Some years later, some of the members of this club form a club with the same name, and the same rules. We ask: 'Have these people reconvened the *very same* club? Or have they merely started up *another* club, which is exactly similar?' There might be an answer to this question. The original club might have had a rule explaining how, after such a period of non-existence, it could be reconvened. Or it might have had a rule preventing this. But suppose that there is no such rule, and no legal facts, supporting either answer to our question. And suppose that the people involved, if they asked our question, would not give it an answer. There would then be no answer to our question. The claim 'This is the same club' would be *neither true nor false*.

Though there is no answer to our question, there may be nothing that we do not know. This is because the existence of a club is not separate from the existence of its members, acting together in certain ways. The continued existence of a club just involves its members having meetings, that are conducted according to the club's rules. If we know all the facts about how people held meetings, and about the club's rules, we know everything there is to know. This is why we would not be puzzled when we cannot answer the question, 'Is this the very same club?' We would not be puzzled because, even without answering this question, we can know everything about what happened. If this is true of some question, I call this question *empty*.

When we ask an empty question, there is only one fact or outcome that we are considering. Different answers to our question are merely different descriptions of this fact or outcome. This is why, without answering this empty question, we can know everything that there is to know. In my example we can ask, 'Is this the very same club, or is it merely another club, that is exactly similar?' But these are not here two different possibilities, one of which must be true.

When an empty question has no answer, we can decide to *give* it an answer. We could decide to call the later club the same as the original club. Or we could decide to call it another club, that is exactly similar. This is not a decision between different views about what really happened. Before making our decision, we already knew what happened. We are merely choosing one of two different descriptions of the very same course of events.

If we are Reductionists about personal identity, we should make similar claims. We can describe cases where, between me now and some future person, the physical and psychological connections hold only to reduced degrees. If I imagine myself in such a case, I can always ask, 'Am I about to die? Will the resulting person be me?' On the Reductionist View, in some cases there would be no answer. My question would be *empty*. The claim that I was about to die would be neither true nor false. If I knew the facts about both physical continuity and psychological connectedness, I would know everything there was to know. I would know everything, even though I did not know whether I was about to die, or would go on living for many years.

When it is applied to ourselves, this Reductionist claim is hard to believe. In such imagined cases, something unusual is about to happen. But most of us are inclined to believe that, in any conceivable case, the question 'Am I about to die?' must have an answer. And we are inclined to believe that this answer must be either, and quite simply, Yes or No. Any future person must be either me, or someone else. These beliefs I call the view that *our identity must be determinate*.

I shall next describe two explanatory claims. The first answers a new question. What unites the different experiences that are had by a single person at the same time? While I type this sentence, I am aware of the movements of my fingers, and can see the sunlight on my desk, and can hear the wind ruffling some leaves. What unites these different experiences? Some claim: the fact that they are all *my* experiences. These are the experiences that are being had, at this time, by a particular person, or *subject of experiences*. A similar question covers my whole life. What unites the different experiences that, together, constitute this life? Some give the same answer. What unites all of these experiences is, simply, that they are all mine. These answers I call the view that *psychological unity is explained by ownership*.

The views described so far are about the nature of personal identity. I shall end with a pair of views that are about, not the nature of this identity, but its importance. Consider an ordinary case where, even on any version of the Reductionist View, there are two possible outcomes. In one of the outcomes, I am about to die. In the other outcome I shall live for many years. If these years would be worth living, the second outcome would be better for me. And the difference between these outcomes would be judged to be important on most theories about rationality, and most moral theories. It would have rational and moral significance whether I am about to die, or shall live for many years. What is judged to be important here is whether, during these years, there will be someone living who will *be me*. This is a question about personal identity. On one view, in this kind of case, this is always what is important. I call this the view that *personal identity is what matters*. This is the natural view.

The rival view is that *personal identity is not what matters*. I claim

What matters is Relation R: psychological connectedness and/or continuity, with the right kind of cause.

Since it is more controversial, I add, as a separate claim

In an account of what matters, the right kind of cause could be any cause.

It is in imaginary cases that we can best decide whether what matters is Relation R or personal identity. One example may be the Branch-Line Case, where my life briefly overlaps with that of my Replica. Suppose that we believe that I and my Replica are two different people. I am about to die, but my Replica will live for another forty years. If personal identity is what matters, I should regard my prospect here as being nearly as bad as ordinary death. But if what matters is Relation R, with any cause, I should regard this way of dying as being about as good as ordinary survival.

The disagreement between these views is not confined to imaginary cases. The two views also disagree about all of the actual lives that are lived. The disagreement is here less sharp, because, on both views, all or nearly all these lives contain the relation that matters. On all of the plausible views about the nature of personal identity, personal identity nearly always coincides with psychological continuity, and roughly coincides with psychological connectedness. But, as I shall argue later, it makes a great difference which of these we believe to be what matters. If we cease to believe that our identity is what matters, this may affect some of our emotions, such as our attitude to ageing and to death. And, as I shall argue, we may be led to change our views about both rationality and morality.

I have now given a first description of several different views. Stated in this abstract way, this description cannot be wholly clear. But what is now

obscure may, when I discuss these views, become clear.

How are these views related to each other? I shall claim, what some deny, that many of these views stand or fall together. If this is so, it will be easier to decide what the truth is. When we see how these views are related, we shall find, I believe, that we have only two alternatives. It is worth stating in advance some of the ways in which, as I shall argue, these views are related.

If we do not believe that we are separately existing entities, can we defensibly believe that personal identity is what matters? Some writers think we can. I shall argue that we cannot.

If we do not believe that we are separately existing entities, can we defensibly believe that personal identity does not just consist in physical and psychological continuity, but is a further fact? I believe that we cannot.

If we believe that our identity must be determinate, must we believe that we are separately existing entities? Having the first belief does not imply having the second. We might believe both that we are not separately existing entities, and that, to any question about personal identity, there must always be an answer, which must be either Yes or No. There are some writers who accept this view. But I shall argue that this view is indefensible. Only if we are separately existing entities can it be true that our identity must be determinate.

It would be possible to claim that we are separately existing entities, but deny that our identity must be determinate. But there are few people who would combine these claims.

Suppose next that we believe that psychological unity is explained by ownership. We believe that the unity of a person's consciousness at any time is explained by the fact that this person's different experiences are all being had by this person. And we believe that the unity of a person's whole life is explained by the fact that all of the experiences in this life are had by this person. These are the explanations given by those who claim that we are separately existing entities. Can we give these explanations if we reject that claim? Some writers suggest that we can. But I shall argue that we cannot.

I shall also argue for the following conclusions:

- (1) We are not separately existing entities, apart from our brains and bodies, and various interrelated physical and mental events. Our existence just involves the existence of our brains and bodies, and the doing of our deeds, and the thinking of our thoughts, and the occurrence of certain other physical and mental events. Our identity over time just involves (a) Relation R—psychological connectedness and/or psychological continuity—with the right kind of cause, provided (b) that this relation does not take a 'branching' form, holding between one person and two different future people.
- (2) It is not true that our identity is always determinate. I can always ask, 'Am I about to die?' But it is not true that, in every case, this

question must have an answer, which must be either Yes or No. In some cases this would be an empty question.

- (3) There are two unities to be explained: the unity of consciousness at any time, and the unity of a whole life. These two unities cannot be explained by claiming that different experiences are had by the same person. These unities must be explained by describing the relations between these many experiences, and their relations to this person's brain. And we can refer to these experiences, and fully describe the relations between them, without claiming that these experiences are had by a person.
- (4) Personal identity is not what matters. What fundamentally matters is Relation R, with any cause. This relation is what matters even when, as in a case where one person is R-related to two other people, Relation R does not provide personal identity. Two other relations may have some slight importance: physical continuity, and physical similarity. (In the case of some people, such as those who are very beautiful, physical similarity may have great importance.)

Here is a brief sketch of how I shall argue for my conclusions. I shall first try to answer some objections to my claim that we could describe our lives in an *impersonal* way. I shall then try to show that, even if we are not aware of this, we are naturally inclined to believe that our identity must always be determinate. We are inclined to believe, strongly, that this *must* be so. I shall next argue that this natural belief cannot be true unless we are separately existing entities. I shall then argue for conclusion (1), that we are not such entities. And I shall argue that, because (1) is true, so are my other three conclusions.

Most of us would accept some of the claims that I shall be denying. I shall thus be arguing that most of us have a false view about ourselves, and about our actual lives. If we come to see that this view is false, this may make a difference to our lives.

HOW WE ARE NOT WHAT WE BELIEVE

THE different views about personal identity make different claims about actual people, and ordinary lives. But the difference between these views is clearer when we consider certain imaginary cases. Most of the arguments that I shall discuss appeal, in part, to such cases. It may be impossible for some of these cases to occur, whatever progress may be made in science and technology. I distinguish two kinds of case. Some cases contravene the laws of nature. I call these *deeply* impossible. Other cases are *merely technically* impossible.

Does it matter if some imagined case would never be possible? This depends entirely on our question, or what we are trying to show. Even in science it can be worth considering deeply impossible cases. One example is Einstein's thought-experiment of asking what he would see if he could travel beside some beam of light at the speed of light. As this example shows, we need not restrict ourselves to considering only cases which are possible. But we should bear in mind that, depending on our question, impossibility may make some thought-experiment irrelevant.

I start with an objection to the Psychological Criterion.

80. DOES PSYCHOLOGICAL CONTINUITY PRESUPPOSE PERSONAL IDENTITY?

I remember trying, when a child, to remain standing among the crashing waves of the Atlantic Ocean. I am the same person as the child who had that experience. On Locke's view, what makes me the same person as that child is my memory, or 'consciousness', of that experience.

Bishop Butler thought this a 'wonderful mistake'. It is, he wrote, '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'.¹²

I have already revised Locke's view. The Psychological Criterion appeals, not to single memories, but to the continuity of memory, and, more broadly, to Relation R, which includes other kinds of psychological continuity. But this revision does not answer Butler's objection.

On one interpretation, the objection would be this: 'It is part of our concept of memory that we can remember only *our own* experiences. The continuity of memory therefore presupposes personal identity. The same is therefore true of your Relation R. You claim that personal identity just consists in the holding of Relation R. This must be false if Relation R itself presupposes personal identity.'

To answer this objection, we can define a wider concept, *quasi-memory*. I have an accurate quasi-memory of a past experience if

- (1) I seem to remember having an experience,
 - (2) *someone* did have this experience,
- and
- (3) my apparent memory is causally dependent, in the right kind of way, on that past experience.

On this definition, ordinary memories are a sub-class of quasi-memories. They are quasi-memories of our own past experiences.¹³

We do not quasi-remember other people's past experiences. But we might begin to do so. The causes of long-term memories are memory-traces. It was once thought that these might be localized, involving changes in only a few brain cells. It is now more probable that a particular memory-trace involves changes in a larger number of cells. Suppose that, even if this is true, neuro-surgeons develop ways to create in one brain a copy of a memory-trace in another brain. This might enable us to quasi-remember other people's past experiences.

Consider

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.

What should Jane believe about these apparent memories? Suppose that, because she has seen this church in photographs, she knows it to be San Giorgio, in Venice. She also knows that she has never been to Italy, while Paul goes to Venice often. Since she knows that she has received copies of some of Paul's memory-traces, she could justifiably assume that she may be quasi-remembering some of Paul's experiences in Venice.

Let us add this detail to the case. Jane seems to remember seeing something extraordinary: a flash of lightning coming from the dark cloud, which forked and struck both the bell-tower of San Giorgio and the red

funnel of a tug-boat passing by. She asks Paul whether he remembers seeing such an extraordinary event. He does, and he has kept the issue of the *Gazzettino* where it is reported. Given all of this, Jane should not dismiss her apparent memory as a delusion. She ought to conclude that she has an accurate quasi-memory of how this flash of lightning looked to Paul.

For Jane's quasi-memories to give her knowledge about Paul's experiences, she must know roughly how they have been caused. This is not required in the case of ordinary memories. Apart from this difference, quasi-memories would provide a similar kind of knowledge about other people's past lives. They would provide knowledge of what these lives were like, *from the inside*. When Jane seems to remember walking about the Piazza, hearing the gulls, and seeing the white church, she knows part of what it was like to be Paul, on that day in Venice.

Jane's apparent memories may be, in one respect, mistaken. It may be claimed: 'Since Jane seems to remember *seeing* the lightning, she seems to remember *herself* seeing the lightning. Her apparent memory may tell her accurately what Paul's experience was like, but it tells her, falsely, that it was *she* who had this experience.'

There may be a sense in which this claim is true. Jane's apparent memories may come to her in what Peacocke calls *the first-person mode of presentation*.¹⁴ Thus, when she seems to remember walking across the Piazza, she might seem to remember seeing a child running *towards her*. If this is what she seems to remember, she must be seeming to remember *herself* seeing this child running towards her.

We might deny these claims. In a dream, I can seem to see myself from a point of view *outside* my own body. I might seem to see myself running towards this point of view. Since it is *myself* that I seem to see running in this direction, this direction cannot be towards *myself*. I might say that I seem to see myself running towards *the seer's point of view*. And this could be said to be the direction in which Jane seems to remember seeing this child run. So described, Jane's apparent memory would include no reference to herself.

Though we could deny that Jane's apparent memories must seem, in part, to be about herself, there is no need to do so. Even if her apparent memories are presented in the first-person mode, Jane need not assume that, if they are not delusions, they must be memories of her *own* experiences. Even if she seems to remember herself seeing the forked lightning, she could justifiably conclude that she is quasi-remembering one of Paul's experiences.

Some of Jane's apparent memories would clearly not be of her own experiences. This would be true of an apparent memory of shaving 'her' beard, while seeing Paul's face in the mirror. In the case of other apparent memories, she might have to work out whether it was she or Paul who had some past experience. And this might sometimes be impossible. She might have to say, 'I do vividly seem to remember hearing that tune. But I do not know whether it was I or Paul who heard it.' When Jane's apparent memories

come to her like this, they are in one respect unlike the apparent memories the rest of us have. Because we do not have quasi-memories of other people's past experiences, our apparent memories do not merely come to us in the first-person mode. They come with a belief that, unless they are delusions, they are about our own experiences. But, in the case of experience-memories, this is a separable belief. If like Jane we had quasi-memories of other people's past experiences, these apparent memories would cease to be automatically combined with this belief.¹⁵

Return now to Butler's objection to the Psychological Criterion of personal identity. On this objection, the continuity of memory cannot be, even in part, what makes a series of experiences all the experiences of a single person, since this person's memory presupposes his continued identity.

On the interpretation that I gave above, memory presupposes identity because, on our concept of memory, we can remember only our own past experiences. This objection can now be answered. We can use the wider concept of quasi-memory.

In our statement of our revised Psychological Criterion, we should not claim that, if I have an accurate quasi-memory of some past experience, this makes me the person who had this experience. One person's mental life may include a few quasi-memories of experiences in some other person's life, as in the imagined case of Jane and Paul. Our criterion ignores a few such quasi-memory connections. We appeal instead to overlapping chains of many such connections. My mental life consists of a series of very varied experiences. These include countless quasi-memories of earlier experiences. The connections between these quasi-memories and these earlier experiences overlap like the strands in a rope. There is *strong connectedness* of quasi-memory if, over each day, the number of direct quasi-memory connections is at least half the number in most actual lives. Overlapping strands of strong connectedness provide *continuity of quasi-memory*. Revising Locke, we claim that the unity of each person's life is in part created by this continuity. We are not now appealing to a concept that presupposes personal identity. Since the continuity of quasi-memory does not presuppose personal identity, it may be part of what constitutes personal identity. It may be part of what makes me now and myself at other times one and the same person. (I say 'part' because our criterion also appeals to the other kinds of psychological continuity.)

Butler's objection may be interpreted in a different way. He may have meant: 'In memory we are directly aware of our own identity through time, and aware that this is a separate, further fact, which cannot just consist in physical and psychological continuity. We are aware that each of us is a persisting subject of experiences, a separately existing entity that is not our brain or body. And we are aware that our own continued existence is, simply, the continued existence of this subject of experiences.'

Does our memory tell us this? Are we directly aware of the existence of this separate entity, the subject of experiences? Some have thought that we are aware of this, not just in memory, but in all of our experiences.

81. THE SUBJECT OF EXPERIENCES

Reid writes:

my personal identity. . . implies the continued existence of that indivisible thing that I call myself. Whatever this self may be, it is something which thinks, and deliberates, and resolves, and acts, and suffers. I am not thought, I am not action, I am not feeling; I am something that thinks, and acts, and suffers.¹⁶

In one sense, this is clearly true. Even Reductionists do not deny that people exist. And, on our concept of a person, people are not thoughts and acts. They are thinkers and agents. I am not a series of experiences, but the person who *has* these experiences. A Reductionist can admit that, in this sense, a person is *what has* experiences, or the *subject of experiences*. This is true because of the way in which we talk. What a Reductionist denies is that the subject of experiences is a *separately existing entity*, distinct from a brain and body, and a series of physical and mental events.

Is it true that, in memory, we are directly aware of what the Reductionist denies? Is each of us aware that he is a persisting subject of experiences, a separately existing entity that is not his brain and body? Is each of us aware, for example, that he is a Cartesian Ego?

This is not a point that can be argued. I do not believe that *I* am directly aware that I am such an entity. And I assume that I am not unusual. I believe that no one is directly aware of such a fact.

Suppose that I *was* aware that I was such an entity. There would still be an objection to the Cartesian View. It has been claimed that I could not know that this entity continued to exist. As both Locke and Kant argued,¹⁷ there might be a series of such entities that were psychologically continuous. Memories might be passed from one to the next like a baton in a relay race. So might all other psychological features. Given the resulting psychological continuity, we would not be aware that one of these entities had been replaced by another. We therefore cannot know that such entities continue to exist.

Reconsider the Branch-Line Case, where it is clear that I remain on Earth. It might seem to a certain person that he has just had these two thoughts: 'Snow is falling. So it must be cold.' But the truth might be this. This person is my Replica on Mars. Just before I pressed the green button, I thought 'Snow is falling'. Several minutes later, my Replica suddenly becomes conscious, in a similar cubicle on Mars. When he becomes conscious, he has apparent memories of living my life, and in particular he seems to remember just having thought, 'Snow is falling'. He then thinks 'So it must be cold'. My Replica on Mars would now be in a state of mind

exactly like mine when I have just had both these thoughts. When my Replica is in this state of mind, he would believe that both these thoughts were had by the same thinker, himself. But this would be false. I had the first thought, and my Replica only had the second.

This example is imaginary. But it seems to show that we could not tell, from the content of our experiences, whether we really are aware of the continued existence of a separately existing subject of experiences. The most that we have are states of mind like that of my Replica. My Replica falsely believes that he has just had two thoughts. He is not aware of the continued existence of a separately existing entity: the thinker of these thoughts. He is aware of something less, the psychological continuity between his life and mine. In the same way, when we have had a series of thoughts, the most that we are aware of is the psychological continuity of our stream of consciousness. Some claim that we are aware of the continued existence of separately existing subjects of experiences. As Locke and Kant argued, and our example seems to show, such awareness cannot in fact be distinguished from our awareness of mere psychological continuity. Our experiences give us no reason to believe in the existence of these entities. Unless we have other reasons to believe in their existence, we should reject this belief.

This conclusion is not, as some write, crudely verificationist. I am not assuming that only what we could know could ever be true. My remarks make a different assumption. I am discussing a general claim about the existence of a particular kind of thing. This is claimed to be a separately existing entity, distinct from our brains and bodies. I claim that, if we have no reasons to believe that such entities exist, we should reject this belief. I do not, like verificationists, claim that this belief is senseless. My claim is merely like the claim that, since we have no reason to believe that water-nymphs or unicorns exist, we should reject these beliefs.¹⁸

Even if we are not directly aware of the existence of these entities, some claim that we can deduce their existence from any of our experiences. Descartes, famously, made such a claim. When he asked if there was anything that he could not doubt, his answer was that he could not doubt his own existence. This was revealed in the very act of doubting. And, besides assuming that every thought must have a thinker, Descartes assumed that a thinker must be a Pure Ego, or spiritual substance. A Cartesian Pure Ego is the clearest case of a separately existing entity, distinct from the brain and body.¹⁹

Lichtenberg claimed that, in what he thought to be most certain, Descartes went astray. He should not have claimed that a thinker must be a separately existing entity. His famous *Cogito* did not justify this belief. He should not have claimed, 'I think, therefore I am'. Though this is true, it is misleading. Descartes could have claimed instead, 'It is thought: thinking is going on'. Or he could have claimed, 'This is a thought, therefore at least

one thought is being thought'.²⁰

Because we ascribe thoughts to thinkers, we can truly claim that thinkers exist. But we cannot deduce, from the content of our experiences, that a thinker is a separately existing entity. And, as Lichtenberg suggests, because we are not separately existing entities, we could fully describe our thoughts without claiming that they have thinkers. We could fully describe our experiences, and the connections between them, without claiming that they are had by a subject of experiences. We could give what I call an *impersonal* description.

As I have said, some writers reject *both* this last Reductionist claim *and* the Cartesian View. These writers do not believe in Cartesian Pure Egos. And they do not believe that a person is any other kind of separately existing entity. They believe that the existence of a person just consists in the existence of his brain and body, and the doing of his deeds, and the occurrence of various other physical and mental events. But these writers claim that we cannot refer to particular experiences, or describe the connections between them, unless we refer to the person who has these experiences. On their view, the unity of a mental life cannot be explained in an impersonal way.

Strawson discusses an argument for this view, suggested by Kant. This argument claims that we could not have knowledge of the world about us unless we believe ourselves to be persons, with an awareness of our identity over time. Shoemaker advances a similar argument. If these arguments are correct, they might refute my claim that we could redescribe our lives in an impersonal way. Because these arguments are at a very abstract level, I shall hope to discuss them elsewhere.^{21a}

Williams discusses a simpler objection to the impersonal description.²¹ This objection is aimed at Lichtenberg. As Williams points out, Lichtenberg's suggested substitute for Descartes' *Cogito* need not be wholly impersonal. It need not be, 'It is thought: thinking is going on'. It could be, 'It is thought: I am thinking'. Since the subject of experiences is here mentioned only in the *content* of the thought, this sentence does not ascribe this thought to a thinker.

Williams then points out that, if several thoughts were expressed in this way, it would need to be made clear whether these thoughts occurred within the same or different lives. This would not be clear if all these thoughts began with the phrase 'It is thought: . . .'. He considers '(T10) It is thought at place A: . . .', but rejects this phrase. He continues:

. . . some less figurative replacement is needed for 'at place A' in the statement of the thought's occurrence—and it is natural to conclude that nothing less than a personal name, or some such, will do as a replacement, so that T10 will give way to

(T11) A thinks: . . .

At this point... the programme of introducing impersonal formulations... will have finally collapsed.

Williams suggests the answer to this objection. As he writes, 'There might possibly be some replacement for the figurative "places" which served the purposes of effective relativization, but did not go so far as introducing a subject who thinks'. There are many such replacements. Two might be:

In the particular life that contains the thinking of the thought that is expressed by the utterance of this sentence, it is thought:...

or

In the particular life that is now directly causally dependent on body A, it is thought:...

Lichtenberg would then need to explain the unity of a person's life in an impersonal way. He could first revise our concept of quasi-memory. He could claim that an apparent memory is an accurate quasi-memory if

- (1) the apparent memory is of a certain past experience,
- (2) this experience occurred,

and

- (3) the apparent memory is causally dependent, in the right kind of way, on this experience.

He would have to show that the right kind of cause can be described in a way that does not presuppose personal identity. He could then appeal to the other kinds of psychological continuity, such as that which holds between the forming of an intention and the later act in which this intention is carried out. I have yet to show that these other continuities, and their causes, can be described in ways that do not presuppose personal identity. Since they can be so described, as I show in Section 89, they could also be described in an impersonal way. Persons must be mentioned in describing the *content* of countless thoughts, desires, and other experiences. But, as Williams points out, such descriptions do not claim that these experiences are *had* by persons. And, without making this claim, we could describe the interrelations between all of the mental and physical events that together constitute a particular person's life.

Lichtenberg's objection to Descartes thus survives. We can refer to and describe different thoughts, and describe the relations between them, without ascribing these thoughts to thinkers. We do in fact ascribe thoughts to thinkers. Because we talk in this way, Descartes could truly claim, 'I think, therefore I am'. But Descartes did not show that a thinker must be a separately existing entity, distinct from a brain and body, and various mental and physical events.²²

82. HOW A NON-REDUCTIONIST VIEW MIGHT HAVE BEEN TRUE

Some writers claim that the concept of a Cartesian Ego is unintelligible. I doubt this claim. And I believe that there might have been evidence supporting the Cartesian View.

There might, for example, have been evidence supporting the belief in reincarnation. One such piece of evidence might be this. A Japanese woman might claim to remember living a life as a Celtic hunter and warrior in the Bronze Age. On the basis of her apparent memories she might make many predictions which could be checked by archaeologists. Thus she might claim to remember having a bronze bracelet, shaped like two fighting dragons. And she might claim that she remembers burying this bracelet beside some particular megalith, just before the battle in which she was killed. Archaeologists might now find just such a bracelet buried in this spot, and their instruments might show that the earth had not here been disturbed for at least 2,000 years. This Japanese woman might make many other such predictions, all of which are verified.

Suppose next that there are countless other cases in which people alive today claim to remember living certain past lives, and provide similar predictions that are all verified. This becomes true of most of the people in the world's population. If there was enough such evidence, and there was no other way in which we could explain how most of us could know such detailed facts about the distant past, we might have to concede that we have accurate quasi-memories about these past lives. We might have to conclude that the Japanese woman has a way of knowing about the life of a Celtic Bronze Age warrior which is like her memory of her own life.

It might next be discovered that there is no physical continuity between the Celtic warrior and the Japanese woman. We might therefore have to abandon the belief that the carrier of memory is the brain. We might have to assume that the cause of these quasi-memories is something purely mental. We might have to assume that there is some purely mental entity, which was in some way involved in the life of the Celtic warrior, and is now in some way involved in the life of the Japanese woman, and which has continued to exist during the thousands of years that separate the lives of these two people. A Cartesian Ego is just such an entity. If there was sufficient evidence of reincarnation, we might have reason to believe that there really are such entities. And we might then reasonably conclude that such an entity is what each of us really is.

This kind of evidence would not directly support the claim that Cartesian Egos have the other special properties in which Cartesians believe. Thus it would not show that the continued existence of these Egos is all-or-nothing. But there might have been evidence to support this claim. There might have been various kinds or degrees of damage to a person's brain which did not in any fundamental way alter this person, while other kinds or degrees of

damage seemed to produce a completely new person, in no way psychologically continuous with the original person. Something similar might have been true of the various kinds of mental illness. We might have generally reached the conclusion that these kinds of interference either did nothing at all to destroy psychological continuity, or destroyed it completely. It might have proved impossible to find, or to produce, intermediate cases, in which psychological connectedness held to reduced degrees.

Have we good evidence for the belief in reincarnation? And have we evidence to believe that psychological continuity depends chiefly, not on the continuity of the brain, but on the continuity of some other entity, which either exists unimpaired, or does not exist at all? We do not in fact have the kind of evidence described above. Even if we can understand the concept of a Cartesian Pure Ego, or spiritual substance, we do not have evidence to believe that such entities exist. Nor do we have evidence to believe that a person is any other kind of separately existing entity. And we have much evidence both to believe that the carrier of psychological continuity is the brain, and to believe that psychological connectedness could hold to any reduced degree.²³

I have conceded that the best-known version of the Non-Reductionist View, which claims that we are Cartesian Egos, may make sense. And I have suggested that, if the facts had been very different, there might have been sufficient evidence to justify belief in this view. Some who believe in Cartesian Egos do not connect them, in such ways, to observable facts. They accept the possibility described by Locke and Kant. On their view, the Cartesian Ego that I am might suddenly cease to exist and be replaced by another Ego. This new Ego might 'inherit' all of my psychological characteristics, as in a relay race. On this *Featureless Cartesian View*, while you are reading this page of text, you might suddenly cease to exist, and your body be taken over by some new person who is merely exactly like you. If this happened, no one would notice any difference. There would never be any evidence, public or private, showing whether or not this happens, and, if so, how often. We therefore cannot even claim that it is unlikely to happen. And there are other possibilities. On this view, history might have gone just as it did, except that I was Napoleon and he was me. This is not the claim that Derek Parfit might have been Napoleon. The claim is rather that I am one Cartesian Ego, and that Napoleon was another, and that these two Egos might have 'occupied' each other's places.²⁴

When the belief in Cartesian Egos is in this way cut loose from any connections with either publicly observable or privately introspectible facts, the charge that it is unintelligible becomes more plausible. And it is not clear that Cartesians can avoid this version of their view. It is not clear that they can deny the possibility described by Locke and Kant. But it is enough to repeat that we have sufficient reasons to reject this view.

83. WILLIAMS'S ARGUMENT AGAINST THE PSYCHOLOGICAL CRITERION

I have defended the Psychological Criterion in two ways. I have claimed, and partly shown, that we can describe psychological continuity in a way that does not presuppose personal identity. And I have claimed that, on the evidence we have, the carrier of this continuity is not an entity that exists separately from a person's brain and body.

I shall next consider another objection to the Psychological Criterion. This is advanced by Williams.²⁵ This objection seems to show that, if some person's brain continues to exist, and to support consciousness, this person will continue to exist, however great the breaks are in the psychological continuity of this person's mental life.

Here is a simpler version of this objection. Consider

Williams's Example. I am the prisoner of some callous neuro-surgeon, who intends to disrupt my psychological continuity by tampering with my brain. I shall be conscious while he operates, and in pain. I therefore dread what is coming.

The surgeon tells me that, while I am in pain, he will do several things. He will first activate some neurodes that will give me amnesia. I shall suddenly lose all of my memories of my life up to the start of my pain. Does this give me less reason to dread what is coming? Can I assume that, when the surgeon flips this switch, my pain will suddenly cease? Surely not. The pain might so occupy my mind that I would even fail to notice the loss of all these memories.

The surgeon next tells me that, while I am still in pain, he will later flip another switch, that will cause me to believe that I am Napoleon, and will give me apparent memories of Napoleon's life. Can I assume that this will cause my pain to cease? The natural answer is again No. To support this answer, we can again suppose that my pain will prevent me from noticing anything. I shall not notice my coming to believe that I am Napoleon, and my acquiring a whole new set of apparent memories. When the surgeon flips this second switch, there will be no change at all in what I am conscious of. The changes will be purely dispositional. It will only become true that, if my pain ceased, so that I could think, I would answer the question 'Who are you?' with the name 'Napoleon'. Similarly, if my pain ceased, I would then start to have delusory apparent memories, such as those of reviewing the Imperial Guard, or of weeping with frustration at the catastrophe of 1812. If it is only such changes in my dispositions that would be brought about by the flipping of the second switch, I would have no reason to expect this to cause my pain to cease.

The surgeon then tells me that, during my ordeal, he will later flip a third switch, that will change my character so that it becomes just like Napoleon's. Once again, I seem to have no reason to expect the

flipping of this switch to end my pain. It might at most bring some relief, if Napoleon's character, compared with mine, involved more fortitude.

In this imagined case, nothing that I am told seems to give me a reason to expect that, during my ordeal, I shall cease to exist. I seem to have as much reason to dread all of the pain. This reason does not seem to be removed by the other things I have to dread—losing my memories, and going mad, becoming like, and thinking that I am, Napoleon. As Williams claims, this argument seems to show that I can have reason to fear future pain whatever psychological changes precede this pain. Even after all these changes, it will be I who feels this pain. If this is so, the Psychological Criterion of personal identity is mistaken. In this imagined case, between me now and myself after the ordeal, there would be no continuity of memory, character, and the like. What is involved in my continuing to exist therefore cannot be such continuity.²⁶

It may be objected that, if I remain conscious throughout this ordeal, there will at least be one kind of psychological continuity. Though I lose all my memories of my past life, I would have memories of my ordeal. In particular, I would continue to have short-term memories of the last few moments, or what is sometimes called the *specious present*. Throughout my ordeal there would be an overlapping chain of such memories.

To meet this objection we can add one feature to the case. After I have lost all my other memories, I am for a moment made unconscious. When I regain consciousness, I have *no* memories. As the ordeal continues, I would have new memories. But there would be no continuity of memory over my moment of unconsciousness.

It may next be objected that I have described this story in question-begging terms. Thus I suggested that, when I am made to lose my memories, I might, because of my pain, fail to notice any change. This description assumes that, after the loss of my memories, the person in pain would still be me. Perhaps the truth is that, at this point, I would cease to exist, and a new person would start to exist in my body.

Williams would reply that, even though my description assumes that I would continue to exist, this is the overwhelmingly plausible assumption. It is the defender of the Psychological Criterion who must show that this assumption is not justified. And this would be hard to show. It is hard to believe that, if I was made to lose my memories while I was in agony, this would cause me to cease to exist half-way through the agony. And it is hard to believe that the change in my character would have this effect.

Williams's argument seems to refute the Psychological Criterion. It seems to show that the true view is the Physical Criterion. On this view, if some person's brain and body continue to exist, and to support consciousness, this person will continue to exist, however great the breaks are in the psychological continuity of this person's mental life.

84. THE PSYCHOLOGICAL SPECTRUM

I shall now revise Williams's argument. Why this is worth doing will emerge later.

Williams discusses a single case in which, after a few changes, there will be no psychological continuity. I shall discuss a *spectrum*, or range of cases, each of which is very similar to its neighbours. These cases involve all of the possible degrees of psychological connectedness. I call this the *Psychological Spectrum*.

In the case at the far end, the surgeon would cause very many switches to be simultaneously flipped. This would cause there to be no psychological connections between me and the resulting person. This person would be wholly like Napoleon.

In the cases at the near end, the surgeon would cause to be flipped only a few switches. If he flipped only the first switch, this would merely cause me to lose a few memories, and to have a few apparent memories that fit the life of Napoleon. If he flipped the first two switches, I would merely lose a few more memories, and have a few more of these new apparent memories. Only if he flipped all of the switches would I lose all my memories, and have a complete set of Napoleonic delusions.

Similar claims are true about the changes in my character. Any particular switch would cause only a small change. Thus, if I am to be like Napoleon, I must become more bad-tempered, and must cease to be upset by the sight of people being killed. These would be the only changes produced by the flipping of the first two switches.

In this revised version of the argument, which involves very many different cases, we must decide which are the cases in which I would survive. In the case at the near end, the surgeon does nothing. In the second case, I would merely lose a few memories, have a few delusions, and become more bad-tempered. It is clear that, in this case, I would survive. In the third case, the changes would be only slightly greater. And this is true of any two neighbouring cases in this range. It is hard to believe both that I would survive in one of these cases, and that, in the next case, I would cease to exist. Whether I continue to exist cannot be plausibly thought to depend on whether I would lose just a few more memories, and have a few more delusory memories, and have my character changed in some small way. If no such small change could cause me to cease to exist, I would continue to exist in all of these cases. I would continue to exist even in the case at the far end of this spectrum. In this case, between me now and the resulting person, there would be *no* psychological connections.

It may be objected:

In this revised form, the argument suspiciously resembles those that are involved in the *Sorites Problem*, or the *Paradox of the Heap*. We are led there, by what seem innocent steps, to absurd conclusions. Perhaps the same is happening here.

Suppose we claim that the removal of a single grain cannot change a heap of sand into something that is not a heap. Someone starts with a heap of sand, which he removes grain by grain. Our claim forces us to admit that, after every change, we still have a heap, even when the number of grains becomes three, two, and one. But we know that we have reached a false conclusion. One grain is not a heap.

In your appeal to the Psychological Spectrum, you claim that no small change could cause you to cease to exist. By making enough small changes, the surgeon could cause the resulting person to be in no way psychologically connected with you. The argument forced you to conclude that the resulting person would be you. This conclusion may be just as false as the conclusion about the grain of sand.

To defend this version of Williams's argument, I need not solve the Sorites Problem. It will be enough to make the following remarks.

When considering heaps, we all believe that there are borderline cases. Are two grains of sand a heap, or four, or eight, or sixteen? We may not know how to answer all of these questions. But we do not believe that this is the result of ignorance. We do not believe that each of these questions must have an answer. We know that the concept of a heap is vague, with vague borderlines. And when the Sorites Argument is applied to heaps, we are happy to solve the problem with a *stipulation*: an arbitrary decision about how to use the word 'heap'. We might decide that we shall not call nine grains a heap, but we shall call heaps any collection of ten or more grains. We shall then be abandoning one premise of the argument. On our new more precise concept, the removal of a single grain may turn a heap of sand into something that is not a heap. This happens with the removal of the tenth last grain.

When it is applied to other subjects, such as phenomenal colour, the Sorites Argument cannot be so easily dismissed.²⁷ Nor does this dismissal seem plausible when the argument is applied to personal identity. Most of us believe that our own continued existence is, in several ways, unlike the continued existence of a heap of sand.

Reconsider the range of cases in the Psychological Spectrum. Like Williams's Example, these cases provide an argument against the Psychological Criterion. This criterion is one version of the Reductionist View. A Reductionist might say:

The argument assumes that, in each of these cases, the resulting person either would or would not be me. This is not so. The resulting person would be me in the first few cases. In the last case he would not be me. In many of the intervening cases, neither answer would be true. I can always ask, 'Am I about to die? Will there be some person living

who will be me?' But, in the cases in the middle of this Spectrum, there is no answer to this question.

Though there is no answer to this question, I could know exactly what will happen. This question is, here, *empty*. In each of these cases I could know to what degree I would be psychologically connected with the resulting person. And I could know which particular connections would or would not hold. If I knew these facts, I would know everything. I can still ask whether the resulting person would be *me*, or would merely be *someone else* who is partly like me. In some cases, these are two different possibilities, one of which must be true. But, in *these* cases, these are not two different possibilities. They are merely two descriptions of the very same course of events.

These remarks are analogous to remarks that we accept when applied to heaps. We do not believe that any collection of sand must either be, or not be, a heap. We know that there are borderline cases, where there is no obvious answer to the question 'Is this still a heap?' But we do not believe that, in these cases, there must *be* an answer, which must be either Yes or No. We believe that, in these cases, this is an empty question. Even without answering the question, we know everything.

As Williams claims, when applied to our own existence, such remarks seem incredible. Suppose that I am about to undergo an operation in the middle of this Spectrum. I know that the resulting person will be in agony. If I do not know whether or not I shall be the person in agony, and I do not even know whether I shall still be alive, how can I believe that I *do* know exactly what will happen? I do not know the answer to the most important questions. It is very hard to believe that these are empty questions.

Most of us believe that we are not like heaps, because our identity must be determinate. We believe that, even in such 'borderline cases', the question 'Am I about to die?' must have an answer. And, as Williams claims, we believe that the answer must be either, and quite simply, Yes or No. If someone will be alive, and will be suffering agony, this person either will or will not be me. One of these must be true. And we cannot make sense of any third alternative, such as that the person in agony will be *partly* me. I can imagine being only partly in agony, because I am drifting in and out of consciousness. But if someone will be fully conscious of the agony, this person cannot be partly me.

The Reductionist View would provide an answer to Williams's argument. When Williams gives his version of this argument, he rejects this view. He concludes instead that, if my brain continues to exist, and to be the brain of a living person, I shall be that person. This would be so even if, between myself now and myself later, there would be *no* psychological connections. After advancing his argument, Williams writes that this conclusion may 'perhaps' be wrong, 'but we need to be shown what is wrong with it'.²⁸

85. THE PHYSICAL SPECTRUM

One objection is that a similar argument applies to physical continuity. Consider another range of possible cases: the *Physical Spectrum*. These cases involve all of the different possible degrees of physical continuity.

In the case at the near end of this spectrum, there would later be a person who would be fully continuous with me as I am now, both physically and psychologically. In the case at the far end, there would later be a person who would be psychologically but not physically continuous with me as I am now. The far end is like the case of Teletransportation. The near end is the normal case of continued existence.

In a case close to the near end, scientists would replace 1% of the cells in my brain and body with exact duplicates. In the case in the middle of the spectrum, they would replace 50%. In a case near the far end, they would replace 99%, leaving only 1% of my original brain and body. At the far end, the 'replacement' would involve the complete destruction of my brain and body, and the creation out of new organic matter of a Replica of me.

What is important in this last case is not just that my Replica's brain and body would be entirely composed of new matter. As I explained, this might become true in a way that does not destroy my brain and body. It could become true if there is a long series of small changes in the matter in my body, during which my brain and body continue to exist, and to function normally. This would be like the ship that becomes entirely composed of new bits of wood, after fifty years of piecemeal repairs. In both of these cases, the complete change in the identity of the components does not disrupt physical continuity. Things are different in the case at the far end of the Physical Spectrum. There is here no physical continuity, since my brain and body are completely destroyed, and it is only later that the scientists create, out of new matter, my Replica.

The first few cases in this range are now believed to be technically possible. Portions of brain-tissue have been successfully transplanted from one mammal's brain to another's. And what is transplanted could be a part of the brain that, in all individuals, is sufficiently similar. This could enable surgeons to provide functioning replacements for some damaged parts of the brain. These actual transplants proved to be easier than the more familiar transplants of the kidney or the heart, since a brain seems not to 'reject' transplanted tissue in the way in which the body rejects transplanted organs.²⁹ Though the first few cases in this range are even now possible, most of the cases will remain impossible. But this impossibility will be merely technical. Since I use these cases only to discover what we believe, this impossibility does not matter.

Suppose we believe that, at the far end of this spectrum, my Replica would not be me. He would merely be someone else who was exactly like me. At the near end of this spectrum, where there would be no replacement, the resulting

person would be me. What should I expect if what will happen is some intermediate case? If they replaced only 1%, would I cease to exist? This is not plausible, since I do not need all of my brain and body. But what about the cases where they would replace 10%, or 30%, or 60%, or 90%?

This range of cases challenges the Physical Criterion, which is another version of the Reductionist View. Imagine that you are about to undergo one of these operations. You might try to believe this version of Reductionism. You might say to yourself:

In any central case in this range, the question 'Am I about to die?' has no answer. But I know just what will happen. A certain percentage of my brain and body will be replaced with exact duplicates of the existing cells. The resulting person will be psychologically continuous with me as I am now. This is all there is to know. I do not know whether the resulting person will be me, or will be someone else who is merely exactly like me. But this is not, here, a real question, which must have an answer. It does not describe two different possibilities, one of which must be true. It is here an empty question. There is not a real difference here between the resulting person's being *me*, and his being *someone else*. This is why, even though I do not know whether I am about to die, I know everything.

I believe that, for those who accept the Physical Criterion, this is the right reaction to this range of cases. But most of us would not yet accept such claims.

If we do not yet accept the Reductionist View, and continue to believe that our identity must be determinate, what should we claim about these cases? If we continue to assume that my Replica would not be me, we are forced to the following conclusion. There must be some critical percentage which is such that, if the surgeons replace less than this percent, it will be me who wakes up, but if they replace more than this percent, it will *not* be me, but only someone else who is merely like me. We might suggest a variant of this conclusion. Perhaps there is some crucial part of my brain which is such that, if the surgeons do not replace this part, the resulting person will be me, but if they do, it will be someone else. But this makes no difference. What if they replace different percentages of this crucial part of my brain? We are again forced to the view that there must be some critical percentage.

Such a view is not incoherent. But it is hard to believe. And something else is true, that makes it even harder to believe. We could not *discover* what the critical percentage is, by carrying out some of the cases in this imagined spectrum. I might say, 'Try replacing 50% of the cells in my brain and body, and I shall tell you what happens'. But we know in advance that, in every case, the resulting person would be inclined to believe that he is me. And this would not show that he *is* me. Carrying out such cases could not provide the answer to our question.

These remarks assume that all of a person's psychological features depend upon the states of the cells in his brain and nervous system. I assume that an organic Replica of me would be psychologically exactly like me. If we reject this assumption, we could respond to this range of imagined cases in a different way. I answer this response in the next section.

If my assumption is correct, and each of these resulting people would be exactly like me, what should we believe about this range of cases? We have three alternatives

- (1) We could accept the Reductionist reply given above.
- (2) We could believe that there is a sharp borderline between two cases. If the surgeons replaced only certain cells, the resulting person would be me. If instead they replaced just a few more cells, the resulting person would not be me, but would merely be exactly like me. There must be this sharp borderline somewhere in this range of cases, even though we could never discover where this line is.
- (3) We could believe that, in all of these cases, the resulting person would be me.

Of these three conclusions, (3) seems to most people the least implausible. If we accept (3), we believe that psychological continuity provides personal identity. We believe that this is so even when this continuity does not have its normal cause: the continued existence of a particular brain.

When we considered the Psychological Spectrum, Williams's argument seemed to show that psychological continuity is not necessary for personal identity. Physical continuity would be sufficient. When we consider the Physical Spectrum, a similar argument seems to show that physical continuity is not necessary for personal identity. Psychological continuity would be sufficient.

We could accept both of these conclusions. We could claim that either kind of continuity provides personal identity. Though this hybrid view is coherent, it is open to grave objections. One objection arises if we combine, not our two conclusions, but the two arguments for these conclusions.

86. THE COMBINED SPECTRUM

Consider another range of possible cases. These involve all of the possible variations in the degrees of *both* physical *and* psychological connectedness. This is the *Combined Spectrum*.

At the near end of this spectrum is the normal case in which a future person would be fully continuous with me as I am now, both physically and psychologically. This person would be me in just the way that, in my actual life, it will be me who wakes up tomorrow. At the far end of this spectrum

the resulting person would have no continuity with me as I am now, either physically or psychologically. In this case the scientists would destroy my brain and body, and then create, out of new organic matter, a perfect Replica of someone else. Let us suppose this person to be, not Napoleon, but Greta Garbo. We can suppose that, when Garbo was 30, a group of scientists recorded the states of all the cells in her brain and body.

In the first case in this spectrum, at the near end, nothing would be done. In the second case, a few of the cells in my brain and body would be replaced. The new cells would *not* be exact duplicates. As a result, there would be somewhat less psychological connectedness between me and the person who wakes up. This person would not have all of my memories, and his character would be in one way unlike mine. He would have some apparent memories of Greta Garbo's life, and have one of Garbo's characteristics. Unlike me, he would enjoy acting. His body would also be in one way less like mine, and more like Garbo's. His eyes would be more like Garbo's eyes. Further along the spectrum, a larger percentage of my cells would be replaced, again with dissimilar cells. The resulting person would be in fewer ways psychologically connected with me, and in more ways connected with Garbo, as she was at the age of 30. And there would be similar changes in this person's body. Near the far end, most of my cells would be replaced with dissimilar cells. The person who wakes up would have only a few of the cells in my original brain and body, and between her and me there would be only a few psychological connections. She would have a few apparent memories that fit my past, and a few of my habits and desires. But in every other way she would be, both physically and psychologically, just like Greta Garbo.

These cases provide, I believe, a strong argument for the Reductionist View. The argument again assumes that our psychological features depend upon the states of our brains. Suppose that the cause of psychological continuity was not the continued existence of the brain, but the continued existence of a separately existing entity, like a Cartesian Ego. We could then claim that, if we carried out such operations, the results would *not* be as I have described them. We would find that, if we replaced much of someone's brain, even with dissimilar cells, the resulting person would be exactly like the original person. But there would be some critical percentage, or some critical part of the brain, whose replacement would utterly destroy psychological continuity. In one of the cases in this range, the carrier of continuity would cease either to exist, or to interact with this brain. The resulting person would be psychologically totally unlike the original person.

If we had reasons to believe this view, it would provide an answer to my argument. There *would* be, in this range of cases, a sharp borderline. And *this* borderline *could* be discovered. It would correspond with what appeared to be a complete change in personal identity. This view would also explain how the replacement of a few cells could totally destroy psychological continuity.

And this view could be applied to both the Psychological and the Physical Spectrum. We could claim that, in both these Spectra, the results would not in fact be what I assumed.

Except for the cases close to the near end, the cases in the Combined Spectrum are, and are likely to remain, technically impossible. We therefore cannot directly discover whether the results would be as I assumed, or would instead be of the kind just described. But what the results would be depends on what the relation is between the states of someone's brain and this person's mental life. Have we evidence to believe that psychological continuity depends chiefly, not on the continuity of the brain, but on the continuity of some other entity, which either exists unimpaired, or does not exist at all? We do not in fact have the kind of evidence that I described above. And we have much reason to believe both that the carrier of psychological continuity is the brain, and that psychological connectedness could hold to any reduced degree.

Since our psychological features depend on the states of our brains, these imagined cases are only technically impossible. If we could carry out these operations, the results would be what I have described. What should we believe about the different cases in this Combined Spectrum? Which are the cases in which I would continue to exist?

As before, we could not find the answer by actually performing, on me and other people, operations of the kind imagined. We already know that, somewhere along the Spectrum, there would be the first case in which the resulting person would believe that he or she was not me. And we have no reason to trust this belief. In this kind of case, who someone is cannot be shown by who he thinks he is. Since experiments would not help, we must try to decide now what we believe about these cases.

In considering the first two Spectra, we had three alternatives: accepting a Reductionist reply, believing that there must be some sharp borderline, and believing that the resulting person would in every case be me. Of these three, the third seemed the least implausible conclusion.

In considering the Combined Spectrum, we cannot accept this conclusion. In the case at the far end, the scientists destroy my brain and body, and then make, out of new matter, a Replica of Greta Garbo. There would be no connection, of any kind, between me and this resulting person. It could not be clearer that, in this case, the resulting person would *not* be me. We are forced to choose between the other two alternatives.

We might continue to believe that our identity must be determinate. We might continue to believe that, to the question 'Would the resulting person be me?', there must always be an answer, which must be either and quite simply Yes or No. We would then be forced to accept the following claims:

Somewhere in this Spectrum, there is a sharp borderline. There must be

some critical set of the cells replaced, and some critical degree of psychological change, which would make all the difference. If the surgeons replace slightly fewer than these cells, and produce one fewer psychological change, it will be me who wakes up. If they replace the few extra cells, and produce one more psychological change, I shall cease to exist, and the person waking up will be someone else. There must be such a pair of cases somewhere in this Spectrum, *even though there could never be any evidence where these cases are.*

These claims are hard to believe. It is hard to believe (1) that the difference between life and death could just consist in any of the very small differences described above. We are inclined to believe that there is *always* a difference between some future person's being me, and his being someone else. And we are inclined to believe that this is a *deep* difference. But between neighbouring cases in this Spectrum the differences are trivial. It is therefore hard to believe that, in one of these cases, the resulting person would quite straightforwardly be me, and that, in the next case, he would quite straightforwardly be someone else.

It is also hard to believe (2) that there must be such a sharp borderline, somewhere in the Spectrum, though we could never have any evidence where the borderline would be. Some would claim that, if there could never be such evidence, it makes no sense to claim that there must somewhere be such a line.

Even if (2) makes sense, claims (1) and (2), taken together, are extremely implausible. I believe that they are even more implausible than the only other possible conclusion, which is the Reductionist View. We should therefore now conclude that the Reductionist View is true. On this view, in the central cases of the Combined Spectrum, it would be an empty question whether the resulting person would be me. This Spectrum provides, as I claimed, a strong argument for this view.

There are some people who believe that our identity must be determinate, though they do not believe that we are separately existing entities, distinct from our brains and bodies, and our experiences. This view I believe to be indefensible. What explains the alleged fact that personal identity is always determinate? The answer must be that the true criterion of personal identity covers every case. The true criterion must draw a sharp borderline somewhere in the Combined Spectrum. But, if we are not separately existing entities, how could there be such a borderline? What could make it true that, in one case, the resulting person would be me, and in the next he would not be me? What would the difference consist in?

There are other people who believe that, though we are not separately existing entities, personal identity is a further fact. These people believe that personal identity does not just consist in the different kinds of

physical and psychological continuity. This is another view that I believe to be indefensible. If we are not separately existing entities, in what could this further fact consist? What could make this fact, in the cases in this range, either hold or fail to hold?

This Spectrum shows, I believe, that certain views must be held together. We cannot defensibly believe that our identity involves a further fact, unless we also believe that we are separately existing entities, distinct from our brains and bodies. And we cannot defensibly believe that our identity must be determinate, unless we believe that the existence of these separate entities must be all-or-nothing.

Some people believe that the identity of *everything* must always be determinate. These people accept a strict form of the doctrine *no entity without identity*. This is the claim that we cannot refer to a particular object, or name this object, unless our criterion of identity for this object yields a definite answer in every conceivable case. On this view, we often mistakenly believe that we are referring to some object, when, because there is no such criterion of identity, there is no such object. It would thus be claimed that most of us mistakenly believe that the name 'France' refers to a nation. On this view, nations cannot be referred to, since they do not exist. There is no criterion of identity for nations which meets the required standard—which would tell us, in every conceivable case, whether or not some nation had continued to exist. Those who hold this view may believe that it could not be similarly true that persons do not exist. If this view is true, and persons do exist, the criterion of personal identity must yield a definite answer in all cases.

This view need not involve the belief that a person is a separately existing entity. This may seem to make this view more plausible. But, if we hold this view, we must again believe that the true criterion of personal identity draws a sharp borderline, quite unknowably, somewhere in the Combined Spectrum. As I have claimed, if personal identity does not involve a further fact, it is very hard to believe that there can be such a line. This is even less plausible than the Reductionist View.

There is another way in which some writers claim that our identity must be determinate. On this view, we have inconsistent beliefs if there are cases where we cannot answer a question about the identity of some object. I believe that there are such cases, and that in such a case the identity of some object is indeterminate. I claim that, in such a case, the statement 'This is the same object that we had before' would be neither true nor false. It has been argued that this claim is incoherent.³¹ I believe that this argument has been answered.³² But suppose that the argument is correct. This implies the following. When we find cases that are not covered by what we believe to be some criterion of identity, we should revise our beliefs by extending this criterion. If we hold this view, we do not believe that the true criterion of personal identity must draw some sharp borderline somewhere in the Combined Spectrum. Rather we believe that, to avoid incoherence, we should draw such a line.

This view hardly differs from the Reductionist View. If we do draw such a line, we cannot believe that it has, intrinsically, either rational or moral significance. We must pick some point on this Spectrum, up to which we will call the resulting person me, and beyond which we will call him someone else. Our choice of this point will have to be arbitrary. We must draw this line between two neighbouring cases, though the difference between them is, in itself, trivial. If this is what we do, this should not affect our attitude towards these two cases. It would be clearly irrational for me to regard the first case as being as good as ordinary survival, while regarding the second case as being as bad as ordinary death. When I consider this range of cases, I naturally ask, 'Will the resulting person be me?' By drawing our line, we have chosen to *give* an answer to this question. But, since our choice was arbitrary, it cannot justify any claim about what matters. If this is how we answer the question about my identity, we have made it true that, in this range of cases, personal identity is *not* what matters. And this is the most important claim in the Reductionist View. Our view differs only trivially from this view. Reductionists claim that, in some cases, questions about personal identity are indeterminate. We add the claim that, in such cases, we ought to give these questions answers, even if we have to do so in a way that is arbitrary, and that deprives our answers of any significance. I regard this view as one version of Reductionism, the tidy-minded version that abolishes indeterminacy with uninteresting stipulative definitions. Since the difference is so slight, I shall ignore this version of this view.

On the simplest version of Physicalism, every mental event is an event in a brain. I remarked above that we could both be Physicalists and accept the *Psychological* Criterion of personal identity. I should add that Reductionists need not be Physicalists. If we are not Physicalists, we could be either Dualists, who believe that mental events are different from physical events, or Idealists, who believe that all events are purely mental. If we believe that we are Cartesian Egos, we believe in one form of Dualism. But Dualists can be Reductionists about personal identity. We can believe that mental events are distinct from physical events, and believe that the unity of a person's life just consists in the various kinds of connection which hold between all of the mental and physical events which, together, constitute this life. This is the Dualistic version of the Reductionist View.

I shall argue that, if we are Reductionists, we should not try to decide between the different criteria of personal identity. One reason is that personal identity is not what matters. Before I argue for this conclusion, I shall explain further what a Reductionist claims. And since most of us are strongly inclined to reject these claims, considering the Combined Spectrum may not be enough to change our view. I shall therefore, in the next chapter, advance other arguments for the Reductionist View.

Reductionists admit that there is a difference between numerical identity and exact similarity. In some cases, there would be a real difference between some

person's being me, and his being someone else who is merely exactly like me. Many people assume that there must *always* be such a difference.

In the case of nations, or clubs, such an assumption is false. Two clubs could exist at the same time, and be, apart from their membership, exactly similar. If I am a member of one of these clubs, and you claim also to be a member, I might ask, 'Are you a member of the very same club of which I am a member? Or are you merely a member of the other club, that is exactly similar?' This is not an empty question, since it describes two different possibilities. But, though there are two possibilities in a case in which the two clubs co-exist, there may not be two such possibilities when we are discussing the relation between some presently existing club and some past club. There were not two possibilities in the case that I described in Section 79. In this case there was nothing that would justify either the claim that we have the very same club, or the claim that we have a new club that is merely exactly similar. In this case these would *not* be two different possibilities.

In the same way, there are some cases where there is a real difference between someone's being me, and his being someone else who is exactly like me. This may be so in the Branch-Line Case, the version of Teletransportation where the Scanner does not destroy my brain and body. In the Branch-Line Case, my life overlaps with the life of my Replica on Mars. Given this overlap, we may conclude that we are two different people—that we are qualitatively but not numerically identical. If I am the person on Earth, and my Replica on Mars now exists, it makes a difference whether some pain will be felt by me, or will instead be felt by my Replica. This is a real difference in what will happen.

If we return to Simple Teletransportation, where there is no overlap between my life and that of my Replica, things are different. We could say here that my Replica will be me, or we could instead say that he will merely be someone else who is exactly like me. But we should not regard these as competing hypotheses about what will happen. For these to be competing hypotheses, my continued existence must involve a *further fact*. If my continued existence merely involves physical and psychological continuity, we know just what happens in this case. There will be some future person who will be physically exactly like me, and who will be fully psychologically continuous with me. This psychological continuity will have a reliable cause, the transmission of my blueprint. But this continuity will not have its normal cause, since this future person will not be physically continuous with me. This is a full description of the facts. There is no further fact about which we are ignorant. If personal identity does not involve a further fact, we should not believe that there are here two different possibilities: that my Replica will be me, or that he will be someone else who is merely like me. What could make these different possibilities? In what could the difference consist?

Some non-Reductionists would agree that, in this case, there are not two possibilities. These people believe that, in the case of Teletransportation, my Replica would not be me. I shall later discuss a plausible argument for this conclusion. If we would be wrong to say that my Replica is me, the remarks that I have just made apply instead to the central cases in the Physical Spectrum. My Replica might have a quarter of the existing cells in my brain and body, or half, or three-quarters. In these cases there are not two different possibilities: that my Replica is me, or that he is someone else who is merely like me. These are merely different descriptions of the same outcome.

If we believe that there is always a real difference between some person's being me and his being someone else, we must believe that this difference comes somewhere in this range of cases. There must be a sharp borderline, though we could never know where this is. As I have claimed, this belief is even more implausible than the Reductionist View.

In the case of clubs, though there is sometimes a difference between numerical identity and exact similarity, there is sometimes no difference. The question, 'Is it the same, or merely exactly similar?' is sometimes empty. This could be true of people, too. It would be true either at the end or in the middle of the Physical Spectrum.

It is hard to believe that this could be true. When I imagine myself about to press the green button, it is hard to believe that there is not a real question whether I am about to die, or shall instead wake up again on Mars. But, as I have argued, this belief cannot be justified unless personal identity involves a further fact. And there could not be such a fact unless I am a separately existing entity, apart from my brain and body. One such entity is a Cartesian Ego. As I have claimed, there is no evidence in favour of this view, and much evidence against it.