

This collection
analogue
philosophy
and human
the collection
analogue
ancient philosophy
comparative
modern philosophy
approach
which reveals
person, a
be distinctive
essential
essays on
related questions
in ancient
we should
human
This is
extended
the relationship
thinking
who are
implications
history of

Christoph
at the University

Cover illustration

THE PERSON AND THE HUMAN MIND

Issues in Ancient and Modern
Philosophy

Edited by

CHRISTOPHER GILL

CLARENDON PRESS · OXFORD

1990

This collection of analogues and human analogues the collected ancient parallels modern approaches which reveal person, a being distinguished essential essays or related questions in ancient times we should humanly extend this thinking the relationship who are implicated in history

Oxford University Press, Walton Street, Oxford OX2 6DP
Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Petaling Jaya Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland
and associated companies in
Berlin Ibadan

Oxford is a trade mark of Oxford University Press
Published in the United States
by Oxford University Press, New York

© In this collection Oxford University Press 1990

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, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of Oxford University Press

British Library Cataloguing in Publication Data
The person and the human mind—issues in ancient and modern philosophy.

- 1. Man—Philosophical perspectives
I. Gill, Christopher
128
ISBN 0-19-824460-6

Library of Congress Cataloging in Publication Data
The Person and the human mind: issues in ancient modern philosophy/
edited by Christopher Gill.
Bibliography. Includes index.
Includes papers from a conference held in Aberystwyth in July 1986.

- 1. Philosophical anthropology. 2. Intellect. 3. Philosophical anthropology—History. 4. Intellect—History. I. Gill, Christopher John, 1946-
BD450.P46218 1990 126—d.20 89-8675
ISBN 0-19-824460-6

Set by Cambrian Typesetters, Frintley, Surrey
Printed in Great Britain by
Courier International Ltd,
Tiptree, Essex

Preface

This volume of essays originated in a conference held at Aberystwyth in July 1986 on 'Persons and Human Beings: Issues in Ancient and Modern Philosophy'. Seven of the essays are based on papers given at the conference; the remaining essays were either written specifically for this volume, or revised for inclusion in it. With one exception, the essays have not been previously published in any form.¹ The aim of this volume, as of the conference, is to explore analogous issues in classical and modern philosophy relating to the concepts of person and human being. It does not attempt to present a history of the development of these concepts from classical to modern times. The volume is divided into two parts, centred on two related themes: the status and interrelationship of the concepts of person and human being, and the extent to which we should understand ourselves as, essentially, human or rational beings. This division does not reflect the arrangement of papers at the conference; it was adopted in the course of preparing the essays for publication, as a way of underlining certain common concerns in the papers given at the conference and in those subsequently added to the volume.

The planning of the conference and the volume was very much a collaborative process, and I have drawn freely on the advice of my fellow-contributors. I am especially grateful to Adam Morton, Amélie Rorty, and Peter Smith for helping me to explore the modern side of the debate, and to my colleagues in ancient philosophy for their willingness to adopt this particular line of approach to our subject. At the time of the composition of the volume, four of the contributors were members of the Departments of Classics or Philosophy at the University College of Wales, Aberystwyth; I am pleased that the volume can record a period in the life of the College when such cooperation was possible.

Thanks are due to the British Academy and the University of Wales for financial assistance to the conference, and to the University College of Wales, Aberystwyth, for contributing to the cost of typing the volume. David Jackson kindly supplied the

¹ Amélie Rorty's chapter is reprinted, with minor alterations, with the permission of Beacon Press, Boston.

This col
analogo
philosop
and hun
the coll
analogo
ancient
compar
modern
approval
which i
person
be disti
essenti
related
in anci
we shc
human
This
exten
the re
thinki
who a
impli
histor
Chris
at the
Cover

that there is no such thing as the concept of personhood, that there are only highly regionalized functions that seemed, erroneously, to be subsumable in a structured concept. Or we might conclude that the various functions of the concept are sometimes at odds, that the concept of a person cannot function to provide decision procedures for resolving conflicts among competing claims for rights and obligations because it embeds and expresses just those conflicts. Nothing hangs on the choice between these conclusions because neither political practice nor philosophic theory is affected by the outcome. For all practical, and theoretical purposes it doesn't matter whether the concept of a person has multiple and sometimes conflicting functions, or whether there is no single foundational concept that can be characterized as *the* concept of a person. As long as we recognize that such appeals are, in the classical and unpejorative sense of that term, rhetorical, we can continue to appeal to conceptions of persons in arguing for extending political rights, or limiting the exercise of political power. The success of such rhetorical appeals depends on whether the proposed concept expresses some of the active values and practices of the audience. Another metaphysical longing remains unsatisfied. But of course that does not mean that we shall be freed of metaphysical longing, nor even of this particular metaphysical longing.⁶

⁶ This chapter is based on a paper given at the Aberystwyth conference on 'Persons and Human Beings'. An earlier version of this chapter appeared as 'Persons as Rhetorical Categories', *Social Research*, 54 (1987), 55-72; and the present version appears in 'Persons and Personae', in A. O. Rorty, *Mind in Action* (Boston, 1988), 27-46; it forms part of 'Relativism, Persons and Practices' in M. Krausz (ed.), *Relativism: Interpretation and Confrontation* (Notre Dame, Ind., 1989).

2 Why there is no Concept of a Person

Adam Morton

The Aim

It is easy to imagine thinking creatures that are not biologically human. Every human culture has imagined them: gods and demons, ogres and sprites, benevolent and malicious spirits, angels, centaurs and hounhnhms, trolls. They are particularly familiar from Greek and Roman mythology. And it is very natural to use the ease with which we can imagine such creatures as a way of making plausible two conclusions, which are attractive in their own right: (1) that our concept of a thinking and acting being (of a mind that is an agent) potentially covers far more than just biological human beings—if in fact our species is the only species in the universe with the relevant characteristics then this is a sort of a cosmic accident; (2) that the bundle of characteristics that make up a thinking agent defines a single characteristic that is of basic moral importance—so that whatever value attaches to human beings and human life could potentially attach to creatures that were not biologically human. To accept these two conclusions is to accept that there is a concept of a person.

I accept many consequences of these two conclusions. I do not think it at all impossible that we will eventually meet intelligent extraterrestrials whose lives we will think of as morally on a par with our own. And I think that 'species chauvinism', the view that human life is valuable just because it is human, is a moral mistake. But I have come to doubt some of the usual grounds for them. What I doubt is that there is a single set of characteristics which would

Drafts of this paper were read at Aberystwyth, Oxford, and Bristol. Audiences at all three gave very constructive abuse. I am particularly grateful to Steven Blackstock, Troels Engberg-Pedersen, Christopher Gill, Peter Smith, Willie Watts Miller and Christopher Williams.

qualify a creature for intellectual and moral personhood. I think that we overestimate the simplicity and unity of the criteria we would apply in judging a creature's application for personhood, and I think that there are many more unclassifiable cases, in which there is no fact of the matter about whether or to what extent the creature in question is a person. (To put it differently: in the wake of a close encounter with extraterrestrials, it is likely that rather than being able to say that they are persons or non-persons or somewhere manageably on the boundary between, we may find ourselves at a loss for words, unable to describe in anything like the person/non-person distinction what we have discovered.) It is not just that I think the notion is vague, with wide fuzzy areas around many different edges. Its defenders would admit that. Rather, I suspect that there is no single concept there at all.

(I think that something similar holds for all of a family of related words, in particular 'sentient' and 'alive'. In each case the full wide range of what the world does and can contain may have some shocks for those who think that they have a simple grasp of what the word involves.)

The reason why we can easily give the concept of a person a definiteness that it does not really have is simple. We import into it more biologically parochial characteristics of human beings than we realize. The purpose of this paper is to show how one class of human characteristics, those centring on our possession of a conception of self, can invisibly infiltrate what purports to be a more general conception of a person. I shall first outline what has come to be the standard picture of personhood among modern philosophers, which is in some ways very different from the way the concept emerges in ancient philosophy.¹ Then I shall give two very different examples, each turning on a different aspect of the way in which thinkers conceive of themselves, which are meant to undercut the standard picture.

Will I then have shown that there is just no concept of a person? I certainly will not have shown that the term 'person', when used, say, to raise doubts about the moral status of new-born human

¹ Before the Aberystwyth conference I thought that the situation was simple but paradoxical: classical culture has more, and more serious, examples of non-human persons, than ours, but classical philosophy strangely fails to separate 'person' from 'human'. At the conference I came to realize that this is too simple. See the papers of Troels Engberg-Pedersen and Stephen Clark in this vol.

babies or to argue for the moral importance of adult elephants, is unintelligible. What I will have shown is that our use of the word is based on a set of rather indefinite family resemblances more than on simple and definite criteria, and that the resemblances focus on parochial features of the human organism. It is this parochiality that seems to me the most important point, for one of the main purposes of speaking of personhood, from Kant to Frankfurt and the recent debates about the status of new-born babies and higher animals, has been to separate our conception of what is valuable about ourselves from the details of our biology. If what I am arguing is right, the idea of personhood can only take us a small way into this project.

The Standard Picture

One very natural and plausible approach to describing a concept of a person-in-general is worked out by Harry Frankfurt in a famous paper.² Frankfurt's theory can be seen as exploiting the insight that two very natural aspects of personhood can easily be developed together. One is the idea that persons can have more complicated thoughts than non-persons (cf. Wittgenstein's observation that a dog can think that his master is behind the door, but not that his master will return next Thursday). The other is the idea that a person has a reflexive capacity: a person cannot be completely ignorant about his own thoughts and motivation. Frankfurt's strategy is to use the first idea, in its modern appearance as the doctrine that belief and desire are propositional attitudes, to incorporate some of the consequences of the inevitably more obscure second idea. I shall give a very brief exposition of Frankfurt along these lines.

A belief or a desire always has a content, which we may take to be given by a proposition: John believes that *the earth is round*, Mary wants John to ignore her (i.e. Mary wants that *John ignores Mary*), Sam wants an apple (i.e. Sam wants that *Sam eats an apple*). This content proposition can be complicated, and a natural index

² See 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy*, 68 (1971), 5-20, repr. in G. Wilson (ed.), *Free Will* (Oxford, 1982), 81-110; D. C. Dennett, 'Conditions of Personhood', in A. O. Rorty (ed.), *The Identities of Persons* (Berkeley, 1976), 175-96; repr. in Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology* (Hassocks, Sussex, 1979), 267-85.

of the intellectual capacity of a creature is the complexity of the propositions that can intelligibly be attributed to it. (So that although cats and infants can have desires, it makes no sense to suppose that a cat or two-year-old is wanting that her mother go away in two days time and return three days after that. The concepts of mother and of a day, and of going and returning, may be available to her, but the complexity of the proposition that combines them is beyond her resources.) One very important source of complexity lies in the possibility that the content proposition have another proposition imbedded in it. Thus in 'Mary believes that Sam wants to steal her Volkswagen' the proposition giving the object of Mary's belief is 'Sam wants to steal Mary's Volkswagen', but that proposition itself has a proposition embedded in it, as content of Sam's desire, namely 'Sam steals Mary's Volkswagen'. Any belief or desire whose content refers in this way to another belief and desire is in the wide sense second order. Sometimes the belief or desire that is the object of the 'outside' belief or desire is a belief or desire of the same person, and these are second-order beliefs or desires in the narrow sense. Examples are: Sam believes that he wants Mary's Volkswagen (i.e. Sam believes that Sam wants that Sam has Mary's Volkswagen), Mary wants to believe that Sam is telling the truth (i.e. Mary wants that Mary believes that Sam is telling the truth).

For Frankfurt, in order to be a person a creature must have second-order beliefs and desires in the narrow sense. Particularly important are desires about one's own desires, for, as Frankfurt argues, the possibility of conflict between what one wants and what one wants to want makes possible some of the characteristic motivation of the moral agent. And it is clear in a general way that beliefs about one's beliefs and desires are rather like consciousness and that desires about one's beliefs are part of such things as self-deception. So a creature with second-order states of mind is likely to have a lot of mental structure, and to be capable of living the kind of complicated reflective life that a person lives.

Moreover the capacity to form a conception of oneself based on one's past experience and guiding one's future actions is crucial to the distinction between living a life and merely being alive, and can thus be seen to be necessary if a creature is to be of value as an individual. And to treat an individual as having this capacity to make its own life is to apply to it the value of autonomy, central to

many moral theories.³ When you add to this the fact that, given the intrinsic connections between the concepts of belief and desire, it is hard in principle to conceive of a creature with second-order beliefs not having some capacity for second-order desires, and vice versa, the result is that a good deal of what we would intuitively include in personhood can be made to follow from the possession of desires about one's own desires. Frankfurt takes persons to be just those creatures which have second-order desires about their own desires and which have desires about which of their desires should result in action.

I shall not develop Frankfurt's picture beyond this point, since his own writings are very readable and there are a number of good papers commenting on them. For my purposes, the important feature of Frankfurt's analysis is the way in which the question of personhood comes down simply to the propositional content of the agent's beliefs and desires. Nothing more is required for personhood than the capacity to have desires with a propositional content complex enough to represent the agent's own desires and their connection with action.

Argument One: A Concept of Oneself

I shall present examples, in this section and the next, which are meant to show that the capacity to have desires with the right propositional contents cannot establish uncontroversial personhood. No definite condition we could set on the contents of desires—or, for that matter, of beliefs—could give all that would be needed. Neither Frankfurt's second-order volitions nor any variant idea will do the job.

The strategy for constructing the first example is to begin with a sentient non-person and to add capacities to it in such a way that Frankfurt's definition comes to be satisfied, although the resulting creature is not intuitively recognizable as a full or uncontroversial person. The strategy requires that one be able to conceive of a creature of a certain degree of mental complexity, but falling short

³ See M. Tonley, *Abortion and Infanticide* (Oxford, 1983); R. Wollheim, *The Thread of Life* (Cambridge, 1984); I. Kant, *Groundwork of the Metaphysics of Morals* (Berlin, 1785), tr. H. J. Paton (London, 1964), pt. II; K. Graham, *The Battle of Democracy* (Brighton, 1986), ch. 5; R. Lindley, *Autonomy* (London, 1986).

of personhood. I shall assure that many higher mammals qualify: they have beliefs and desires and can think thoughts of some complexity, but are not persons.

Imagine a super-intelligent rabbit, that is, a creature biologically like a rabbit in physical and sensory organization but with vastly increased capacities of memory and inductive reasoning. I shall call this creature the hyper-rabbit. Imagine further that the hyper-rabbit is capable of using human language, either by the application of enormous inductive powers to her observations of human speech or thanks to the presence of an innate disposition to structure observed speech according to the 'right' syntactical forms. None of these capacities are beyond the biological potentialities of non-human mammals. (For all that, creatures like the hyper-rabbit clearly do not exist on this planet, as becomes more evident below. But I argue later in the paper that the example is nearer the condition of actual creatures than one might at first think.) Such a creature would be capable of playing a role in human life, and I imagine her working in an office. The department of education, say, processing reports on the political opinions and statements of intellectuals.

Since the hyper-rabbit can use language, she can use proper names correctly to refer to particular people, places, things, and animals. Animals include hyper-rabbits, and she understands and uses the public name for herself. ('Hype', say.) And indeed documents about her pass across her desk occasionally, and her reports mention herself as well as others. She thus has beliefs about herself, and indeed beliefs about her own beliefs and desires, since her job is to report on intellectuals' opinions. And since her reports contain recommendations, she has desires about her own desires, which she wants to see translated into action. Her reports will contain recommendations such as: Morton and Hype are thoroughly reliable and their work is to be encouraged, but it would be desirable if their interest in pursuing their own work were more thoroughly subordinated to their civic duty.

She seems thus to have second-order desires, and to qualify for personhood. Her claim seems less strong, though, when we consider some features that I have deliberately omitted from her constitution.

First, although she can have beliefs and desires about herself, all that 'about herself' amounts to is that she understands her own name. She understands it in the way in which she understands the

names of others, by having various beliefs which for the most part truly describe her, involving terms whose use traces back causally to her, and which are sufficiently similar to those of others that communication does not get snarled up. But that is not the way in which *we* have concepts of ourselves. The difference can be brought out by considering how a person normally answers a question such as 'where are you?' Imagine that Hype is on the telephone with a colleague who should bring her some data. The colleague asks where she is. To answer it Hype reasons as follows: 'you' in this context is coextensive with "Hype" [*linguistic competence*]; Hype is an employee of the education ministry [*belief about self*] and in fact looking up under H in the ministry directory 'Hype' is listed as room 999 of the Joseph building; and during working hours Hype is found in that room [*belief about self*]; and so the report to give is that room 999 is Hype's location. Now a human person might on occasion deduce their location in roughly that way, but normally a quite different class of considerations would apply. One would first think: *this* body is at this location in space, and so this is where I am. And then one would go on to find a useful description of this location in space. Hype cannot do it that way, for her conception of herself—the meaning she attaches to 'Hype', and to 'I' in suitable contexts, and which fills out her beliefs and desires about herself—is not essentially linked to any indexical: it is tied neither to 'the person associated with *this* body', nor 'the person at *this* location', nor even 'the person thinking *this* thought', in any way in which the indexical *this* requires a demonstrative awareness of one's immediate situation.

The other gap in Hype's constitution is closely related to this. Being basically a rabbit (in much the way in which a human person is basically a lemur) she has a rabbit's conception of self, alongside her sophisticated edifice of second order beliefs and desires. That shows itself in several ways. She has typical rabbit emotions such as fear and lust and parental concern. And these have primitively conceptualized contents: to fear something is to be apprehensive lest it harm *me*, where *me* is tied not to the sophisticated sense of 'Hype' but to a rabbit's bodily awareness. It governs impulses to get her body away from 'here'. And, similarly, she has desires whose content involves the primitive self-concept. If she wants a carrot then she wants that *she* have that carrot. But none of this is conceptualized, and we can simply take all these self-references as

denoting 'here', 'this body', and so on, as the case demands. And so although she has emotions and desires which we might naturally elaborate using an 'I' on her behalf, that 'I' is misleading: it does not represent a concept which enters her thoughts or, in particular, her second-order beliefs and desires.

One result of this is that her second-order desires are not exactly those of ordinary persons. If an ordinary human person wants not to want a carrot the proposition which expresses the content of the second-order want would be

I do not want that I eat a carrot,

where the 'I' represents that indexical description-free reference to self. But if Hype wants not to want a carrot the content of the second order want is

Hype does not want that Hype eats a carrot,

which does not involve the primitive 'I'. As a result when Hype wants not to eat a carrot her second order desire cannot conflict with a first order desire to eat a carrot, whose content proposition would be

I eat a carrot

or

This rabbit eats a carrot

in the direct way in which the human person's second-order desire conflicts with the same first-order desire.

(This point can be put in a slightly different way using a device due to Castaneda. Castaneda would write 'Peter wants to eat a carrot', where the implicit subject of 'eat' is the indexical 'I', as

Peter wants that he* eat a carrot.

Then the logical form of 'Peter wants not to want a carrot' would be expressed as

Peter wants that he* not want that he* eat a carrot.

But the most that can be truly asserted by saying that Hype wants not to want a carrot is the proposition which would be expressed by

Hype wants that Hype does not want that Hype eats a carrot

And the conflict between this and Hype's simple desire for a carrot, expressed in Castaneda style by

Hype wants that she* eat a carrot,

is pretty indirect, since the ultimately embedded propositions in the two cases are not negations of one another.⁴

The essence of the situation is a certain dissociation in Hype's self-concept. There is the descriptive grasp of self involved in her higher-order beliefs and desires, and there is the indexical element involved in the more primitive parts of her psychology, but there is no conception which unifies the two. The result is in some ways rather machine-like and in others rather like a sage or a god: Hype sees herself impersonally as just one individual among others; her idea of herself is made of exactly the same materials as her ideas of others.⁵

One might think that Hype could construct a primitive 'I' for herself. She could answer 'where are you?', for example, by construing the indexical 'you' as 'the individual to whom these words are addressed' and 'I' as 'the individual who is speaking these words', so that she could take the question as asking for an utterance which identified the location of its utterer, and then think out loud 'these words are uttered in room 999, and now think Hype utters after them will have the same location, so an utterance

⁴ See J. Perry, 'The Problem of the Essential Indexical', *Noûs*, 13 (1979), 3-21; id., 'Frege on Demonstratives', *Philosophical Review*, 86 (1977), 474-97; D. Lewis, 'Attitudes De Dicto and De Re', *Philosophical Review*, 87 (1980), 513-43; H. N. Castaneda, 'He*'. A Study in the Logic of Self-Consciousness', *Ratio*, 8 (1966), 130-57.

⁵ More God-like than sage-like, actually, since a Stoic sage, at any rate, has a thoroughly human sense of self, but does not attach much importance to it. See Engberg-Pedersen in this vol., ch. 5. Still, this disregard of something intrinsically human is a more delicate form of Hype's disassociation.

of "I am in room 999" will say, truly, what is asked for.⁶ And it is true that by mastering some standard demonstratives—here, now, this—some of the force of the indexical 'I' can be captured.

But not all of it. There seem to be two sources of trouble. The first is just the imprecision of the demonstratives. For example, agents normally know what they are doing, and Hype too has normal motor control which involves some representation of where her body is and what it is up to (and so, in this sense she does have a primitive representation of self, an 'I') as well as knowing by observation and other means what Hype's current actions are. But she cannot replace 'what I am doing' with 'this action', since there may be too many actions happening in the near vicinity. ("These actions." Which actions? 'Well, we would have to respond, 'these that I'm doing.')

Similarly 'the individual uttering these words' is not really enough to capture 'me the speaker', since others may be speaking at the same time. (Which words? 'Mine, of course.')

There are unanswered questions here about how Hype can keep her first- and second-order states so separate. They are as much questions about human psychology as about Hype. How do ordinary human persons' second-order desires influence their first-order desires? How do desires with a descriptive content influence the largely indexical content of the intentions which shape action? How do our capacities to perceive qualities of things connect with conscious or verbal thought, to become part of the indexical component of conceptualized beliefs and desires? These questions are just as hard and confusing when asked about human beings as when asked about Hype. The last of them is particularly important here. We have more perceptual capacities than we are conscious of, or can put into words, and some of these capacities give content to some of our emotions and inarticulate desires. So we are not so different from Hype in this respect.

The second reason why 'I' cannot be synthesized out of easier demonstratives lies in our capacity for self-knowledge. Human persons know when they are unhappy, in pain, having indigestion, or imagining a holiday in Vienna. The result is that they have propositional beliefs about themselves—"I am unhappy", "I am imagining a holiday in Vienna"—which need some conceptual grasp of the self, but which link it not to public descriptions but to a

⁶ This is essentially Reichenbach's theory of indexical terms. See H. Reichenbach, *Elements of Symbolic Logic* (New York, 1948).

network of innate and acquired capacities of self-attribution. To believe that one is unhappy is to believe that 'this individual is unhappy' where the *this* is, as in the examples above, really a pretty complex business and no improvement on an unexplained 'I'. A human person's self-attributions are often perspectival in this sense—perhaps illusory but a fairly basic part of what it is like to be human. They present themselves as if they were reports on an object, the self, on which the person has a privileged perspective.⁷

Human persons have the status of persons in part because they possess the characteristic grasp of self of human persons. Frankfurt is right to that extent, and it is surely one of the reasons that human persons are both moral agents, having capacities for self-evaluation and normative reflection, and objects of value to other agents. One fact basic to both is that we live lives that develop as unities in accordance with our developing conception of ourselves. But this human grasp of self is no single thing. It is certainly not the same as simply having a name one can apply to the individual who one happens to be. Nor is there any small and natural list of demonstrative capacities in terms of which it can be characterized. What then should we think of an agent who has a grasp of self, but not that characteristically human grasp of self? Hype, for example: is she a person? Two contrasting facts pull us in opposite directions. On the one hand what it is like to be Hype is extremely different from what it is like to be a human person. But on the other hand the life that she leads, the complex of things she can accomplish alone or with others, is not very different from ours. She could be a party to a social contract, and she could be a valued member of a community. I hope that it is clear by now what we should say. If there is no simple unity to the collection of skills that go into using and understanding 'I' and which are presupposed by the normal form of our second-order desires, and beliefs, then such individuals are neither clear persons nor clear non-persons, nor in any useful way in between. They are moral agents in some ways, and in others not, and they are objects of value in some ways and in others not. For moral agency and moral value are not themselves unsplittable conceptual atoms.

⁷ If you need convincing that the indexical grasp of self in introspection is not the same as the indexical awareness of one's body and action, read D. C. Dennett's very entertaining "Where am I?" in *Brainstorms*, pp. 310–23.

Argument Two: Personal Identity

The example I built the first argument around is meant to undermine the synchronic aspect of the concept of a person, that is, it is meant to make you doubt that there is always an objective fact of the matter whether a particular individual at a particular time is a person or not. A number of writers, of whom Derek Parfit is the most successful, have tried to undermine the diachronic aspect of the concept. They have produced examples which are meant to lead you to doubt that it is always an objective matter whether a person at one time is the same as a person at a later time. These examples have a certain fantasy-like or science fictional quality to them. They involve brain transplants, matter transmitters, memory extracting machines, and the like. One might well grant them their intended point and still conclude that the identity-through-time of no *real* person will ever be indeterminate. The aim of this section is to produce an example of a situation which can and does happen to actual human beings, in which their identity is—or so I shall argue—indeterminate.

My cases are inspired by real cases found in neurological literature: people whose memory only extends back a few hours. I shall modify the details in the direction of a neatness which is medically extremely unlikely but not impossible: what I am describing could happen to human beings.⁸

I shall call my person Fred. Fred, as a result of alcoholism and bad luck, is struck by a peculiar form of Korsakoff's psychosis on the first of September 1980. For the rest of his life, his days follow the same general pattern: he wakes in the morning, thinking it must be the second of September 1980. The discovery that it is not is extremely distressing to him, but he is a person of emotional resilience as well as of formidable intellect, and during the day he comes to terms with his situation, to the extent of grasping roughly what it is, and even, on some days, gathering the outline of 'his'

⁸ See H. Gardner, *The Shattered Mind* (New York, 1975), ch. 5; O. Sacks, *The Man Who Mistook his Wife for a Hat* (London, 1985), ch. 2; and, though it presents a rather different case, A. Luria, *The Man with a Shattered World* (London, 1973). On personal identity in general see D. Parfit, *Reasons and Persons* (Oxford, 1984), pt. III. Parfit's purpose is ultimately not to make one doubt the intelligibility of personal identity as much as its importance. S. Shoemaker and R. Swinburne, *Personal Identity* (Oxford, 1984), brings out clearly the centrality of the sense of self-identity in our intuitions about the identities of persons.

history since September 1980. And then in the evening he sleeps, and the memories of that day are swept into oblivion.

Fred's case differs from actual periodic amnesias in the tidy coincidence of the time through which memory is preserved with the sleeping/waking cycle. This coincidence is also found in some fictional variations on the same idea.⁹ It does not affect the issues involved. But it does make the situation much more easily imagined, partly because many of us who habitually wake in a state of disorientation which can take hours to dissipate often find ourselves in a mildly analogous condition. Parfit refers to cases much like these as 'branch line' cases, not dwelling on the fact that they are much more easily realized than many of his other cases.

There are two ways of reacting to cases like Fred's. One of them is implicit in the way I have told the story and obviously comes very naturally. Call it the linear version: there is a person and *he* is terribly damaged and then later *he* spends each day without any memories of the day before. According to the linear version Fred is one person throughout, and Fred at any one time has a radical lack of memories of immediately earlier times. The other way of telling the story comes less easily as a narrative, but makes sense as a description or an analysis. Call it the branching version: there is a person Fred-0 who is terribly damaged, and thereafter there is a succession of persons associated with Fred-0's ageing body. Call them Fred-1, Fred-2, Fred-3, . . . Each Fred-*n* is a continuation of Fred-0, but each Fred-*n* and Fred-*m* are distinct (when *n* and *m* are distinct). According to the branching version Fred is not so much damaged as fragmented into a series of people.

The linear version is the natural way to tell the story. (But then stories are linear things.) It also fits a lot of what would be said. Fred's children going to visit him say 'I'm going to see Fred again. I hope he has got over the cold he had yesterday.' But the branching version also fits a lot of what we would say and do. If Fred on one day makes a promise to do something the next day and then does not, we would not berate him or even remind him of it. The promise is as if made by someone else. In fact, one would avoid receiving promises from Fred, just as one would rather avoid receiving promises from someone who one knows will die before he

⁹ See A. Buddy's, 'The End of Summer', in B. Aldis (ed.), *Penguin Science Fiction* (Harmondsworth, 1965); P. J. Farmer, 'Sketches among the Ruins of my Mind', in H. Harrison (ed.), *Nova 3* (London, 1973), 150-92.

can fulfil them. And if one made a promise to Fred and then found that one could not keep it, one would not apologize to the next day's Fred or feel that one owed *him* some recompense, although one might well feel a guilty debt to the Fred of the day before. Similarly, one would not hold successive Freds responsible for the misdeeds of previous Freds, or in fact assume that any of the moral connections between earlier and later stages of a person's life hold of the succession of Freds.

And to balance the narrative naturalness of the linear version there is the emotional appetitiveness of the branching version. To a friend or relative grieving over the difficulty of maintaining a relationship with someone who does not remember one day what he has done or said the day before one could, and surely at some stage would, argue that the relationship is misdirected: it would be gentler on all concerned to think of the enterprise as a series of encounters with the person one loved and loves, each presenting one with another day's continuation of that person but each disconnected from the others.

Both linear and branching versions can seem right, and there are both physical facts and social intuitions to support each. And each one can to some extent accommodate the intuitions that support the other. The linear version can accommodate the pro-branching intuitions I described just above by appealing to a distinction between what is emotionally convenient to think and what is actually true. And the branching version has two ways of accommodating some of the force of the impulse to say, e.g., that one visited Fred yesterday and will visit him again tomorrow. One way is to construe the pronoun here as one would the 'he' in 'I talked to the Dean of Arts last year, and when the same problem arose this year I went to him again', when it is a different Dean from the one year to the next. And the other is to distinguish between 'same person' and 'same human being': one visited Fred yesterday and will visit the same human being tomorrow, but unfortunately the same human being will not be the same person.

Which one is really right? It should be clear by now, that in the absence of facts or arguments of some completely new kind, the only reasonable course is to accept that both are legitimate ways of seeing the situation. Is Fred-36 the same person as Fred-37? Or, to put the same question in a less natural form, is Fred-37-plus-Fred-

38 (temporal part of) a continuing person? The concept of a person is not up to answering the question. 'Yes', 'No', and 'To some extent', are equally wrong answers; the only right answers begin with 'it depends what you mean by ...'.

Second Case into First Case

The case of fragmented Fred can be made to approach that of heterogeneous Hype, though only by giving it a little of the science fictional quality of the earlier example. Suppose that each morning Fred is first given an understanding of his condition and then a summary of his life since his amnesia. (Imagine that Fred is valued for his unique skills as a government decision-maker and negotiator. Each day there is a new problem that only Fred can handle, but in order to handle it he has to be briefed, and the briefing must include a fair amount of his past history.) Enough information about his past, absorbed thoroughly enough, will give an approximation to the continuity his broken memory-links cannot supply. Extraordinary intellectual power would be needed to manage a good approximation. Postulate the required intellectual power, which would certainly be beyond the ordinary human scale. Is the result—Hyperted—then a single continuing human person?

Again the answer seems to be 'it depends what you mean by ...'. A person is something that lives a life, and a life is something that is knitted together over time in part by the agent's organization of it, her capacity to turn her past into her future.¹⁰ Human persons do this largely by virtue of their ability to remember their pasts. This ability makes a person relatively independent of others for her knowledge of who, historically, she is. And it presents the relevant information in an agent-centred way: a person does not have to sift through all her beliefs about elephants in order to tell whether she rode an elephant last week; she can just remember what she was doing then. Memory information is filed as a special subset of one's information about the world, accessible by different means and presented in a form suitable for making a unity of one's actions. As a result, Hyperted in the modified example will differ from human persons both in his dependence on others (or possibly on

¹⁰ See Wollheim, *Thread of Life*.

mechanical devices) to connect him to 'his' past and in his lack of a distinction between his beliefs about the world-in-general and his beliefs about himself.

Hyperfred and Hype are thus very similar. In fact, since Fred's situation is biologically possible for human beings and since Hyperfred results from Fred simply by the addition of intellectual power, we now have a way of imagining what it is like to be a creature like Hype. First imagine that you lose your memory in the way that Fred has. (I do not think that is very hard to imagine. Terrifying, but not difficult.) And then imagine that each morning you are given and absorb your biography for the time between the onset of your amnesia and the present. (That too is not hard to imagine in principle, just because what one's imagination grasps is the absorbing of vastly less information than would actually be necessary.) Then, after both steps, what one is imagining is the condition of a creature very similar to Hype: all one's second-order beliefs and desires, except those about this day's acts and states, are non-indexical and non-perspectival.

Hyperfred's claims to personhood seem much the same as Hype's. And so if either of them is a person Fred is a potential or quasi-person. And in all three cases it is really the same aspect of human personhood which is lacking: the human capacity to have a concept of self which is linked to a complex of very specific ways of obtaining information about oneself. But one's intuitive willingness to accept Hyperfred and Hype as persons can be rather different. Many people find they want to classify Fred and Hyperfred as persons and exclude Hype. Yet there is not much difference between Hype and Hyperfred when both are described abstractly. The details that make them different leave a lot of room for human sympathy for Fred and Hyperfred, though. Should this be relevant to their moral status? It depends what you are going to use the status for.

Biology versus Imagination?

It may well seem that my arguments so far—and they are the main arguments of the paper—amount to just a quibble about something pretty marginal to our conception of a person. I think that this is not so. Let me try to develop the worry into an objection, so that in

dealing with the objection we can see how fundamental the problems I am raising are.¹¹

To begin, it may seem that the gap I have pointed out in the Frankfurtian concept of a person is easily filled: to the possession of second-order beliefs and desires (and volitions) one needs just to add a requirement that persons have indexical thoughts about their present and past states.

But this cannot be right. Hype has indexical thoughts about her present and past states, inasmuch as she can think about 'this body' and the mental and physical states it supports, and she is not at all obviously a person. And Hyperfred lacks the capacity to have some important indexical thoughts about his past, but it is not at all clear that he is not a person. So just having some indexical thoughts about oneself cannot be enough. They have to be the *right* ones. And how can this be specified except in such question-begging terms as 'the indexical thoughts that normal humans have'?

What is needed is a general specification of the kind of *combination* of indexical capacities that is characteristic of a person's grasp of self. The capacities that normal human beings have keep them in touch with a self which has an immediate spatial location, a past history, a present flow of thought and activity, and a future of projected actions. The capacity for this combination of indexical thoughts is not a luxury for a creature that is going to act in the physical world and carry out tasks that take a long time to finish. And so—one might argue—evolutionary pressures are likely to result in creatures with capacities to co-ordinate direct awareness of such things as their locations, the actions of their bodies, and their planned future actions. Thus even though it may be very hard to say in a sufficiently general way what the right combination of capacities is, it is biologically necessary—according to this line of thought—that there must be a constrained set of right combinations of capacities.

I think there is something right about this form of the objection. It focuses on what a person does rather than on what a person is. It makes the Wollheimian point that some of our primary intuitions concern the kind of life that a person can live, which involves action governed by reasoning, desires governable by reflection, and

¹¹ This objection was put to me at the Aberystwyth conference by Peter Smith and Troels Engberg-Pedersen.

intentions governed by memory, in order to fill out a constantly changing plan. A person is something that can have such a life.

To accept this is not, however, to accept that there is any right combination of capacities which underlies a person's grasp of himself. (Or even any right set of combinations.) For one thing, it is not at all clear that the life of a person requires a particular kind of psychological organization. I do not myself think it impossible for us to discover (or be discovered by) a species clearly living as persons—communicating with each other with something like speech, forming social institutions, creating works of art—but who seem to have a psychology resisting the concepts of belief and desire, let alone of second-order desire and of human-like self-conception. I will not try to argue for such a dramatic conclusion now. In the special case of a person's grasp of self, however, something of the kind is not hard to make plausible.

Hype and Hyperfred are unnatural creations. The argument at hand suggests that creatures like them can only exist in imagination. That is far from obvious. Consider Hyperfred. His characteristic difference from a normal human being is that his contact with his past passes through the assimilation of facts rather than first-person 'from the inside' experiential memory. He never relives his past, but just knows that various things were true in it. Could no natural animal be like this? That is, could a creature not have all its knowledge of its past in terms of factual knowledge rather than first-person memory? Just to ask the question is to wonder what counts as memory. And it does not take philosophical or psychological sophistication to doubt that there is much biological unity to all the things that are commonsensically classified as memory. Human persons typically remember their past in (at least) three distinct ways. They can re-experience events, having replicas or representations of the original experiences. They have a great deal of knowledge of vital identifying and signposting facts about themselves. (Presumably experiential memory would be a confusing jumble without the presence of this knowledge.) And they have an accumulation of skills, learned in the past and reusable to varying degrees. Any great discontinuity in any of these would count as amnesia. In human psychology they are distinct, and no one of them need be a neurological unity.¹² And the grasp that different

¹² See K. V. Wilkes, 'More Brain Lesions', *Philosophy*, 55 (1980), 455–70.

humans have of their pasts may well vary in its dependence on these various different capacities.

The same is true of any of the capacities I mentioned as being employed in a person's life: reasoning, reflection, intention, memory. None of them are inevitable biological unities. And the presence in a creature of the behavioural capacities that they represent in us need not rest on anything like the same complex of psychological functions which give their contribution to what it is like to be a human person.

That is my main defence against the threatening argument, and I derive my final conclusion from it below. Before stating that, though, I think it may help strengthen the intuitive case to consider an example, again of a possible creature, of a quite different sort. This one has *more* indexical capacities intrinsic to its concept of itself than we do. Imagine, adapting a suggestion of Lewis Thomas's,¹³ a dog-like creature with a capacity to tell non-inferentially the degree of genetic relatedness to itself of any member of its species. A creature of this species can locate itself non-inferentially in genetic space and consequently thinks of itself as, among other things 'the individual with *this* genetic constitution', thus relating itself to parents, siblings, and possible descendants. (There would be an obvious biological advantage to this.)

Such a creature might think conceptually, have second-order desires, and live the life of a person together with others of its kind. The capacity I have mentioned could be basic to some part of this life, making it richer than ours, and it might allow the creature to live the life of a person without the use of some capacities which for us are essential. (Perhaps some kinds of human memory might be lacking.) From this creature's point of view creatures like us would lack a very basic element of personhood. A human does not know who he is, in a very basic way. Humans lack a fundamental component of 'I', the capacity to know of a certain node in a genealogy that that is *me*, and as a result the best that human beings can have is a bloodless over-abstract conception of themselves, not tied to any capacity to know which biological individuals they are. Or so my creatures would say.

To end, let me return to the claim with which I began this paper, that we import biologically parochial features into our picture of

¹³ *The Youngest Science* (Oxford, 1985), ch. 19.

personhood. We can now see two ways in which this happens. One is, of course, the assumption that there is a simple unity to our indexical self-grasping skills. The other is the assumption that a creature capable of living a life structured like ours must have a psychology like ours. In fact there are indefinitely many ways both in which a creature can gain the capacity for a richly structured life and in which it can organize its capacity to have thoughts about itself. And if the way in which it does either is sufficiently different to the way in which we do it we find ourselves unable to classify them as persons.

The concept of a person can therefore be developed in either of two directions, and its problems come from the way it is caught between them. They are (1) the capacity to lead the life of a person and (2) the quality of having experience like that of a human person. (The capacity to have human-like indexical thoughts is part of (2): if a creature has the right combination of indexical thoughts, then what it is like to be it will in this respect be like what it is like to be a human person.) Both (1) and (2) capture some important features of personhood. We may think of (1)-type persons (for example Hype) as potential members of social contracts and potential formulators of the categorical imperative. And we may think of (2)-type persons as having a special if mysterious intrinsic value our reverence for which underlies our reluctance to harm or destroy small children or the various Freds, whether or not they are capable of living a planned, extended, contract-making life.

We cannot centre personhood on the (2) aspect, for very few of the creatures whose lives we may come to value and balance against our own will have just our combination of capacities. What it is like to be most of them is just beyond our imaginations. And we cannot centre personhood on the (1) aspect for several reasons. First, this would leave out cases like the Freds which seem intuitively to have a lot of the importance of personhood. Then there is the enormous variety of ways in which an individual in a community may be capable of leading a structured person-like life. Then also there is the indefiniteness of this conception of a person-like life: is it lived by hermits or by the insane or the amoral, let alone by the participants in the more remote social arrangements possible elsewhere in space and time? For all these reasons there really is not a very manageable concept centred on (2) for all that (1)-type considerations would often lead us to treat creatures as persons.

(1)

So there is nothing we can analyse and define and present as 'the concept of a person'. Nothing whose sense will settle in advance the status of all the beings we might value for the reasons we value human persons. For these values are not any single thing. And possible creation is too varied. Not to see this is to think that all advanced creatures must be humans in disguise. To think this is not only to reveal a limited imagination, but also to miss some of the richness of what it is like to be, and how it is good to be, human.