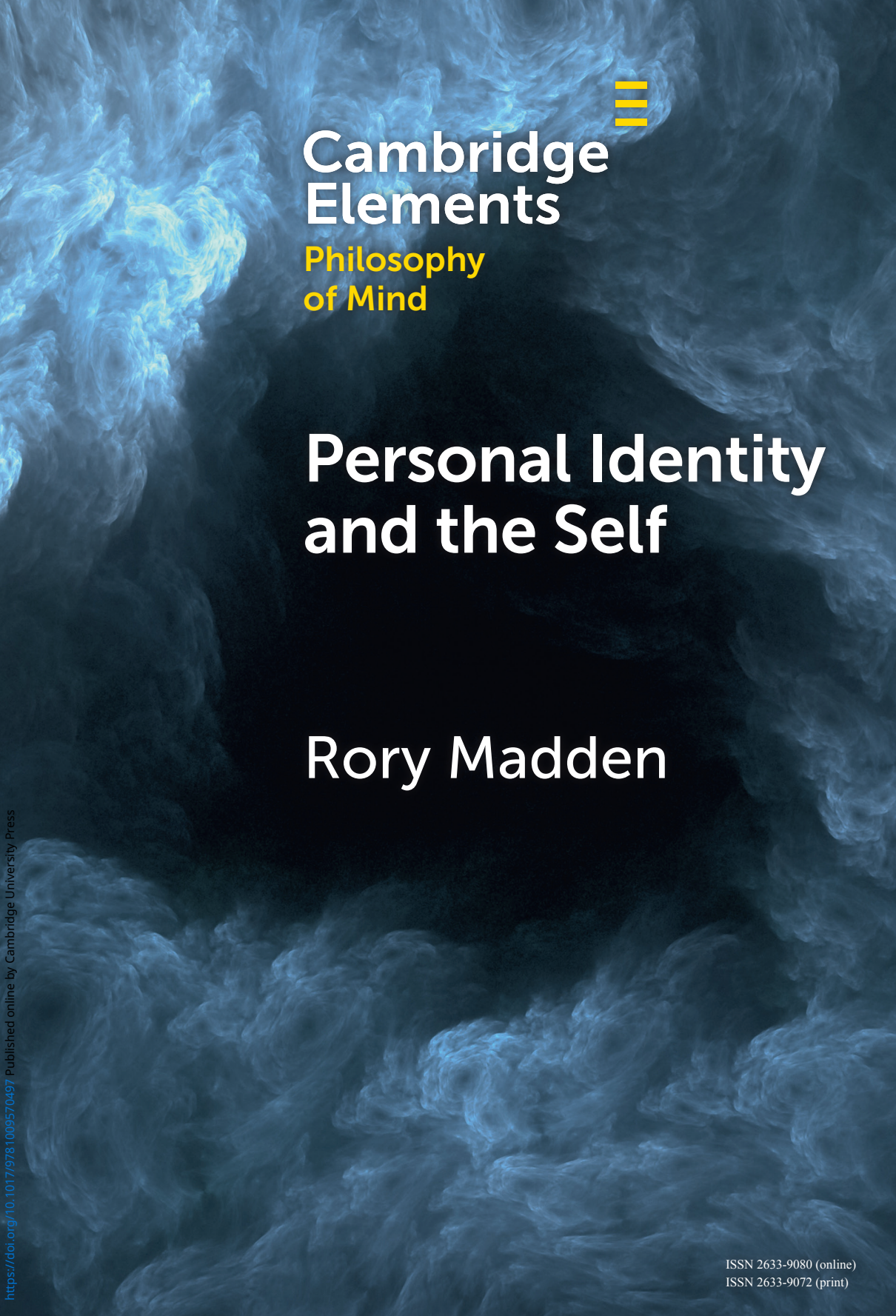


https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

ISSN 2633-9080 (online)
ISSN 2633-9072 (print)

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Cambridge Elements

Elements in the Philosophy of Mind
edited by
Keith Frankish
The University of Sheffield

PERSONAL IDENTITY AND THE SELF

Rory Madden
UCL, London

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781009570527

DOI: 10.1017/9781009570497

© Rory Madden 2024

This publication is in copyright. Subject to statutory exception and to the provisions
of relevant collective licensing agreements, no reproduction of any part may take
place without the written permission of Cambridge University Press & Assessment.

An online version of this work is published at doi.org/10.1017/9781009570497 under
a Creative Commons Open Access license CC-BY-NC 4.0 which permits re-use,
distribution and reproduction in any medium for non-commercial purposes providing
appropriate credit to the original work is given and any changes made are indicated.
To view a copy of this license visit https://creativecommons.org/licenses/by-nc/4.0

When citing this work, please include a reference to the DOI 10.1017/9781009570497

First published 2024

A catalogue record for this publication is available from the British Library.

ISBN 978-1-009-57052-7 Hardback
ISBN 978-1-009-57051-0 Paperback
ISSN 2633-9080 (online)
ISSN 2633-9072 (print)

Cambridge University Press & Assessment has no responsibility for the persistence
or accuracy of URLs for external or third-party internet websites referred to in this
publication and does not guarantee that any content on such websites is, or will
remain, accurate or appropriate.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Personal Identity and the Self

Elements in the Philosophy of Mind

DOI: 10.1017/9781009570497
First published online: December 2024

Rory Madden
UCL, London

Author for correspondence: Rory Madden, r.madden@ucl.ac.uk

Abstract: What are we? What owns our thoughts and experiences? Are we anything at all? After an introduction, Section 2 assesses a 'no-bearer' theory of experience, and the 'no-self' contention that self-representations are about no real entity, before introducing a positive hypothesis about the objects of our self-representations: the 'animalist' claim that we are biological organisms. Section 3 discusses the classic challenge to animalism that brain transplantation is something we could survive but no animal could survive. This challenge introduces positive alternatives to animalism, as well as animalist responses, including one which questions the assumption that psychology is irrelevant to organism persistence. Section 4 surveys a 'thinking parts' problem and conjoined twinning and commissurotomy, also considered problematic for animalism. The interpretation of these cases revisits questions about bearers of experience, objects of self-representation, and the relation of biology and psychology. This title is also available as Open Access on Cambridge Core.

Keywords: personal identity, self, persistence, experience, animalism

© Rory Madden 2024

ISBNs: 9781009570527 (HB), 9781009570510 (PB), 9781009570497 (OC)
ISSNs: 2633-9080 (online), 2633-9072 (print)

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Contents

1 Introduction 1

2 Ownership and Ourselves 5

3 Persistence 28

4 Multiplicity 52

References 74

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

1 Introduction

1.1 'Personal Identity' and 'the Self'

The title will give readers a good idea of the kind of theories and arguments to follow. But we won't be using the expressions 'personal identity' or 'the self' very much. These expressions suggest a potentially confusing variety of different topics.

Locke (1975) popularized the terms 'person' and 'self' in philosophy. He used them both to refer to a being with a reflective perspective on its own mind. He used 'self' to describe such a being from its own reflective perspective. He used 'person' to describe it from the outer perspective of the public forum of law and morality. This neatness of usage has not persisted. 'Person' has roughly kept the sense of a publicly identifiable member of a moral or legal community, but Locke's original connection with self-consciousness has weakened; sometimes 'person' can just mean human being, without commitment to every human being possessing self-consciousness. Elsewhere the legalistic connotation has taken on a life of its own, so that a corporation can be counted a person, again without any commitment to this legal entity being self-conscious, or a human being.

Talk of 'personal identity' inherits the compounding ambiguity of 'identity'. We might be talking about numerically counting persons in one of the above senses. We might be talking about the qualitative sameness of a person, in all respects, or in some privileged respect: we might be talking, as social psychologists do, about a person's 'identity' in the sense of a social role or status of special significance.

Usage of the noun 'self' is even more anarchic. It has vague connotations of reflexivity and deep importance but otherwise slides freely across different categories. Sometimes one's self is simply oneself, the very thing that one is. In other uses one's self is something which is in some sense *in* oneself, an important part or aspect of oneself, and therefore not literally oneself. This ingredient of importance might be one's essence, or character, or whatever is most important about oneself. Here 'self' comes close to one sense of 'personal identity'. In another usage of 'self', the special ingredient of oneself is a supposed inner homuncular agent or experiencer, controlling or witnessing from within oneself.

Another kind of thing 'in' oneself is one's *representation* of oneself. In cognitive science 'self' is sometimes used to talk about one's conception, or experience, of oneself. In a slight loosening of this representational talk which is common among writers at the interface of cognitive science and phenomenology, a 'self' is a notional object one can be said to 'have' just in case one has a representation of oneself. Compare the way in which one can be said to 'have a past' just in case one has memories; or to 'have a world' just in case one has a perceptual perspective. Sometimes 'selves' in this representational sense are multiplied according to distinct ways one has of representing oneself,

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

for example, insofar as one takes oneself to have lived through a course of events, one has a 'narrative self'; insofar as one experiences one's current bodily state, one has a 'bodily self'; and insofar as one simply experiences oneself as experiencing, one has 'a minimal self' (Gallagher 2000). There is also a picturesque way of using the noun 'self' to talk about an aspect of oneself as if it were somebody in its own right. My 1990s self would have hated this music. My professional self has few interests in common with my vacationing self.

1.2 The Question of What We Are

We won't botanize the uses of 'person' and 'self' any further. We will cut a straight path through the jungle and focus on the notion of one's 'self' as the very thing that one is. The current metaphysical debate which is the closest continuer of Locke's seminal discussion of personal identity focuses on a question about the self in this sense. The question is: What am I? Or more commonly, on the reasonable presupposition that readers and writers share a nature, a first-person plural question is asked: What are *we*? The metaphysical debate is about our presupposed shared nature.

Not just any true answer to the question will do. It may be true that we are inhabitants of the solar system. It may be true that we are readers of philosophy. But these answers by themselves would not tell us very much about our basic metaphysical nature. They would not, for example, settle what sort of changes we could undergo while remaining in existence. They would not settle whether we have material parts, or how our parts are unified. Exemplary answers can be found in history of philosophy. Descartes's view was that we are simple immaterial souls. Hume's view was that we are causally interrelated bundles of perceptions. These are views about our basic metaphysical nature. Each would, if true, be a good answer to the question of what we are.

The question: What are we? is what might be called a perspectival question. It belongs with such questions as: What is that? Where did this come from? Who are you? What time is it now? These are perspectival questions in the sense that they use context-dependent expressions like 'that', 'this', 'you', and 'now'. Such expressions make the answer to the question depend upon the particular perspective from which the question is being asked. For example, the answer to a 'what is that?' question depends upon what is in fact being picked out by the pronoun 'that' in the particular context of enquiry. In a different context, words with the same English meaning could be used to ask a question about a different salient object.

'We' is doubly context-dependent. It refers to a plurality which includes whoever is the speaker or thinker of the context. But there is further context-dependence about what precisely is the more inclusive plurality of things.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Is it speaker and hearer? My football team? All sentient beings in the universe? In what follows the 'we' will be understood to include, besides myself, readers whose lives and epistemic situations are similar enough to mine to follow what I'm talking about. I shall assume that we share a basic metaphysical nature. This assumption is not an a priori certainty but it is a reasonable a posteriori assumption as things stand. Artificial intelligence cannot yet understand philosophical writing. Amazon does not yet deliver books to extraterrestrial beings.

The 'we' in question is sometimes glossed by 'human people' or 'human persons' (e.g. Olson 2003; Bailey 2015). A human person is a reflective being which is somehow intimately associated with a body of a human biological character (unlike a possible angelic or divine person). Readers of what follows are unlikely to go wrong understanding 'we' as 'human persons'. But it would be a mistake to suppose that the non-perspectival question: What are human persons? somehow captures the real significance of the perspectival question: What are we? It is widely agreed that non-perspectival knowledge cannot capture the cognitive significance of perspectival knowledge (Perry 1979).

We can illustrate this general point with a thought experiment. Consider the hypothesis that we are in fact Martians in a virtual reality machine which is merely simulating human life on Earth. Perhaps there really are human persons living their lives on Earth, but we are not they. Were we to discover this, the non-perspectival question: What are human persons? would be drained of its distinctive interest. Perhaps not all of its interest: given our long immersion in the machine, the question might retain some sentimental interest, like the concern for soap opera characters of long acquaintance. But the question: What are human persons? grips us in a special way only on the presupposition that *we* are human persons. The question: What are we? expresses a state of curiosity that is ineliminably perspectival.

Philosophical enquiry as an expression of ineliminably perspectival curiosity may seem suspiciously parochial or egocentric. Shouldn't philosophy aim for a God's eye 'view from nowhere'? Philosophers differ in their tastes but it would be a mistake to think that perspectival enquiry is unusual in philosophy: consider Socrates's 'know thyself'. Perspectival questions are not even unusual in metaphysics: it is true that questions like: What are properties? Or: Could there have been nothing? do not appear to be perspectival. But consider the fact that metaphysicians of time enquire into the nature of the present. In a sense, this is enquiry into a subject matter which differs from one perspective of enquiry to the next. But it is far from clear that it should therefore be eliminated in favour of some non-perspectival question. Even if it is presently 1 January 2023, the question: What is the nature of the present? cannot be replaced without loss of significance by the question: What is the nature of 1 January 2023?

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

In other examples of metaphysical enquiry the perspectivality may be more implicit. For example, the perennial question: What are colours? may be a question which can only be asked by a community which has a certain kind of perceptual apparatus. This question may be inaccessible or simply uninteresting to beings with other perspectives. This does not make it a defective philosophical question for us. More pervasively, any philosophical enquiry into the way things in fact are, be it foundational philosophy of physics, or natural language semantics, relies for much of its interest on the presupposition that we are asking about some aspect of *our* world. There may be intellectual amusement to be gained from trying to make consistent sense of the complexities of the laws of magic in some merely possible Dungeons & Dragons universe, or in formalizing the conditional in the merely possible language of High Elvish. But these are not questions about our laws, or our languages.

So we can take the perspectival question: What are we? at face value. It does not seem to be in bad company.

1.3 Overview

Neither the Cartesian nor Humean answer is central to today's debate about what we are. This is not the occasion for a history of the debate but it should be noted that Descartes's and Hume's views were both arrived at 'from the armchair' broadly speaking. Neither rests upon empirical evidence which would be called into question by external world scepticism. Their arguments can perhaps still be said to rest on experience, but only in the thin sense that they take for granted that experience occurs and has a certain general structure, in abstraction from what in particular it appears to be revealing about the world beyond.

Contemporary philosophy of mind and metaphysics is no longer inhibited by this restriction to armchair evidence. Smart's (1959) seminal mind–brain identity theory was an unashamedly a posteriori thesis and Kripke (1980) influentially cleared away general ideological obstacles to accepting that a posteriori investigation could reveal the basic metaphysical natures or essences of things. In accordance with this more naturalistic approach to metaphysics in general, the current debate about what we are takes seriously the view that we are brains and the view that we are biological organisms. Neither of these views is knowable from the armchair.

Other writers find in the messy realities of the natural world support for more negative or sceptical reactions to the question of what we are. Perhaps the cognitive science of self-representation can be argued to show that 'I' fails to pick out anything. Perhaps the awkward empirical facts that there are multiple nested material things in one's vicinity or that there are competing principles for

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

individuating biological objects, show that 'I' fails to pick out anything in particular. Scientific enlightenment may encourage scepticism about any straightforward positive answer.

In what follows we are going to concentrate on broadly naturalistic responses, both positive and negative, to the question of what we are.

Section 2 'Ownership and Ourselves' assesses a 'no-bearer' theory of experience, surveying a variety of forms of ownership of experience, before reviewing some arguments for the 'no-self' contention that our self-representations are about no real entity. These arguments face an uphill struggle. The section ends by introducing a positive theory about the objects of our self-representations, the currently influential animalist theory that we are biological organisms of a certain kind.

This theory is a convenient foil for surveying the contemporary debate about what we are. Section 3 'Persistence' discusses an important challenge to animalism, based on the Lockean contention that there are situations which we could survive but no animal could survive. Brain transplantation is the most widely discussed. The challenge serves to introduce some positive alternatives to animalism, as well as some animalist responses, including a response which questions the common assumption that psychology is irrelevant to the persistence of biological organisms.

Section 4 'Multiplicity' surveys another genre of challenge to animalism, according to which there are more of us at a time than there are animals. Multiplicity cases include a problem of 'thinking parts' and cases of conjoined twinning and commissurotomy also thought to be problematic for animalism. Assessing these challenges require us to revisit questions from Section 2 about what it is to be a bearer of mental phenomena and about the reference of self-representation. Minimal realizer and teleo-functionalist approaches to bearers are introduced. The reader is once again invited to resist the habit of opposing biology and psychology.

2 Ownership and Ourselves

Mental phenomena comprise at least perception, sensation, action, emotion, thought, memory, and planning. These are phenomena distinctively associated with the explanation of animate behaviour, and in some sense also accessible to introspection 'from within'. Such phenomena seem by their very nature to be owned or possessed. Experience requires an experiencer. Thought requires a thinker. We don't think of mental phenomena as free-floating objects like snowflakes or bubbles. Mental phenomena belong to the very things whose animate behaviour they explain, the very things that can access them from within.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

The question of what we are is evidently closely related to questions about mental ownership. What are the thinkers of our thoughts? What are the subjects of our experiences? The almost trivial-sounding answer to these questions is that *we* are the owners of our thoughts and experiences. Answers to the question of what we are, and to the question of what owns our thoughts and experiences, simply march in step.

But the relationship between owners and ourselves turns out to be less straightforward. First, is it really part of the basic nature of mental phenomena to be 'owned'? What does that even mean? Although some philosophers have found the claim obvious, others have dismissed it as obscure ideology and illusion, encouraged by the way we talk but ultimately unsupported by experience. In this section, we will focus on the ownership of sensory experience, reviewing considerations for and against a no-bearer theory of experience. We will then consider the related 'no-self' view that there is nothing that we are.

2.1 Bearers of Experience

Do conscious sensory experiences have 'owners'? As ever in philosophy, one can find oneself pulled in opposing directions.

In Lewis Carrol's *Alice's Adventures in Wonderland*, the grinning Cheshire Cat slowly disappears, leaving behind its grin. No cat: just the grin. The episode is absurd. Insofar as we think of a grin as a thing at all, we think of it as something which is by its nature *of* some entity, some grinner. A grin is *ontologically dependent* on a grinner, where in general x is ontologically dependent on y just in case it is part of the very nature or identity of x that it stands in a certain relation to y.

What is the relation in this case? We might think of a grin like a walk or a fall: an occurrent event in which a certain subject or agent undergoes or does something. Alternatively, we might think of a grin more along the lines of a dent in a car door, or an eddy in a stream: an enduring modification of an entity which exists for just as long as that entity is configured a certain way. Whether we think of grins as undergoings or modifications, each grin is by its nature related to a certain entity. This entity we can call the grin's bearer. Either way, no grin could exist apart from its bearer.

It is a traditional view that experiences should be understood like grins in this respect. An experience is by its very nature *of* some other entity, some experiencer. For each experience, there must be a particular thing to which that experience is by its nature intimately related. How related? Again there are different possibilities. On one way of thinking, an experience is as an event, which must be undergone by something. Another way of thinking of an

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

experience is as a sort of internal dent or eddy which modifies a particular object in a special way. Either way, each experience is by its nature related to something else, which we can call its bearer. Either way, no experience could exist without its bearer.

The recognition of a bearer gains prima facie support from some ordinary language considerations about reference to particular experiences. A natural way to refer to an experience is by 'de-verbalizing' a sentence which is used to say that some subject experiences in a certain way. This is an example of a general way of achieving reference to particular events. For example, from the sentence with a verb 'Tom grinned (at a specified time *t*)' we can form a singular expression which picks out a particular event: 'Tom's grinning at *t*'. Similarly, we can pick out experiences by de-verbalizing sentences which ascribe kinds of experience to subjects at times: from 'Tom smelled a mouse at *t*' we can derive the referring expression 'Tom's smelling a mouse at *t*'; from 'Tom felt pain at *t*' we can derive the referring expression 'Tom's feeling pain at *t*'; and so on. So there is a very basic way of achieving reference to particular experiences which is dependent on the ascription of experience to subjects. It is quite natural to think that this dependence in our way of picking out experiences reflects their ontological dependence.

But pulling in the other direction is a broadly empiricist tradition, in Eastern as well as Western philosophy, which is sceptical about the connection of experiences to bearers. From this point of view, the linguistic considerations are superficial and misleading.

Why? A historical motivation for scepticism about bearers is scepticism about souls. It was commonly presupposed that bearers of experience would have to be indivisible immaterial substances of the sort which feature in, for example, Cartesian and Classical Indian philosophy. But that presupposition is questionable. We will later consider alternatives according to which experiences have bearers which are brains or organisms.

However, we can trace an influential route to scepticism about bearers which does not depend on the presupposition common to substance dualists and their historical opponents, that bearers would have to be souls.

First, in reaction to the ordinary language considerations, the point should be made that the de-verbalization of sentences which ascribe kinds of experience to subjects does not appear to be our *only* way of achieving reference to experiences. It is plausible that we are each in a special position to pick out our own conscious experiences 'from the inside' in virtue of our immediate reflective awareness of them, perhaps by using the emphatic demonstrative expression '*these* experiences'. We don't arrive at these words by derivational morphology from language which refers to subjects which are experiencing. This way of referring seems to be self-standing.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Furthermore, the reflective awareness which enables us to pick out experiences in this direct way does not clearly present them as undergone, as modifications, or in any other way ontologically dependent upon some distinct entity. It's not hard to get into a subjective mood in which it seems that when one reflects on one's conscious sense experience, one is aware of nothing but a succession of sensory quality instances. To illustrate: I sit back in my chair, close my eyes, and try to articulate only what is manifest to unprejudiced reflection on sensory experience. What do I find? There is a fizz and hum of bodily sensation; a reddish-orange visual field with inchoate afterimages flickering and drifting; a distant engine sound against a faint background of tinnitus; and a stale taste of coffee. None of this is given as having a bearer.

Is there any reason to think that absence of evidence in this case is evidence of absence? Here is one possible reason. It is sometimes claimed that conscious experiences are unique among phenomena in that their basic nature is fully revealed to reflective awareness; they have no hidden nature.[1] If this 'revelation' thesis is correct, and if reflective awareness does not reveal conscious experiences to have bearers, then conscious experiences do not have bearers. For if they had bearers then that would be an aspect of their basic nature, and therefore would be something revealed to reflective awareness.

What we will call a *no-bearer theory* endorses this conclusion: conscious experiences do not have bearers. According to this theory, it is not part of the nature of an experience to be related to a certain distinct thing, as, for example, its undergoing or modification.

2.2 Varieties of Ownership

It is important to note that the no-bearer theory can consistently acknowledge a wide variety of other relations to experiences which could reasonably be called 'ownership'. We can follow the no-bearer theorist on a tour of these other ownership relations.

First, I find that my cutaneous, visceral and muscular sensations are all felt to occur within various parts of a single bounded spatial object positioned in a certain way. This bounded object can be agreed to 'have' these bodily sensations in the sense of being their *apparent spatial locus*. It may in the end be true that this locus is one and the same as something which is undergoing conscious events of feeling bodily sensations in its parts. But the identification of locus and bearer would be a non-trivial claim, which goes beyond anything revealed by the inspection of bodily sensation itself. These are different notions of ownership.[2]

[1] Johnston (1992: 223); Chalmers (2003); Strawson (2015); and Goff (2017: 109–110).
[2] See (Brewer 1995) for a contrary point of view.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Next, suppose that I get up from my chair and start exploring my environment. It emerges that the subsequent succession of given sensory quality instances causally depends on the changing state and situation of one particular bounded object, especially upon the location and orientation of its head and eyes, and the positioning of its tactile surfaces. No other object occupies this special place in the causal explanation of the evolution of one's sensory experience. This object then is what might be called the *causal interface* of one's experiences. It is the object upon which one's experiences are especially causally dependent.[3]

The no-bearer theorist need not deny that an experience has a causal owner in this sense. Indeed the no-bearer theorist Schlick (1936) ventured to diagnose our inclination to think that experiences have a bearer as an exaggerated misconstrual of this genuine phenomenon of causal ownership: the causal relation to one object is an intimate and de facto unique relation, and therefore easily mistaken for ontological dependence. But being the causal interface of an experience is not a way of being a bearer of that experience. For it seems we can coherently conceive of a situation in which this experience is caused not by its actual causal interface but by some other object instead: for example, I can conceive of my visual experiences having been brought about by a remote radio link to the eyes of some other body. Causal ownership of experience is contingent or transferable. But metaphysical ownership is not transferable: if it is part of the nature of an experience to be related to a particular thing, then it couldn't possibly fail to be related to that thing.

In an important discussion of persons, P.F. Strawson criticized Schlick's diagnosis of our inclination to believe that experiences have a bearer.[4] He argued that Schlick's diagnosis self-defeatingly presupposes the stronger form of ownership which he was trying to explain away as illusory. Strawson observed that it is not true without qualification that experiences de facto causally depend upon one object. What I find is that *my* experiences de facto causally depend upon one object. It would be bizarrely solipsistic to assume that *all* experiences causally depend on this object. But now if the accurate way to articulate the real causal phenomenon which induces the illusion of a bearer is that *my* experiences causally depend upon a certain object, then the diagnosis is taking for granted that there is a more fundamental way in which experiences can be called 'mine'. What I find, one might think, is that those experiences of which *I am the bearer* are especially causally related to a certain object.

[3] For a recent example of a conception of one's 'self' as simply that which has a special role in the explanation or prediction of one's ongoing sensory experience, see (Hohwy & Michael 2017).

[4] Strawson (1959: 95–98).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Strawson's critique has weaknesses. First, the no-bearer theory is not obviously committed to Schlick's allegedly self-defeating diagnosis of our inclination to believe that experiences have a bearer. The illusion of a bearer might be given some explanation other than the confusion of causal and ontological dependence: for example as resulting from the linguistic fact that we typically achieve reference to experiences through the ascription to subjects of ways of experiencing. Alternatively, the no-bearer theorist might appeal to some non-linguistic illusion. It is sometimes argued that our inclination to believe in a bearer of experience is a side-effect of the particular way in which non-conceptual self-representations are generated by the brain (e.g. Metzinger 2004; Letheby & Gerrans 2017).

A second weakness of Strawson's critique is its presupposition that the further kind of owner presupposed by the use of 'my' must be a bearer. As we shall see shortly, it may be possible to acknowledge a kind of owner which is neither causal interface nor bearer.

Finally, we have already noted that it seems possible to pick out experiences without reference to any owner. One can restrict attention to what are in principle only some among all of the experiences by using the demonstrative 'these experiences', avoiding the possessive 'my' altogether. What I find is that *these* experiences are causally dependent upon this particular body. This way of stating the finding leaves open that there may be other experiences 'out there' which are not dependent on this body.

Still one might press a conceptual problem of solipsism. Unless I think of experiences as inhering in bearers I cannot really make any sense of other experiences being 'out there' at all.[5] Schlick himself insisted that this 'egocentric predicament' is escaped because we can make sense of the possibility that these experiences – what he called 'the data' – could turn out to have multiple causal owners. Imagine from the reflective perspective some of these experiences depend on one object, while others of these experiences depend on some other object: for example, imagine feeling pain when one body is damaged while having visual experiences caused by the ocular stimulation of another body. The problem for Schlick is that it is not clear how what's imagined here could amount to more than a situation in which, as we might put it, one experiencer has two bodies. Just imagining multiply caused experiences from

[5] Frege (1997) argued that recognition that my subjective 'ideas' have a metaphysical bearer makes it possible to conceive of there being other bearers of ideas. Frege was critiquing a subject-less conception of sensations associated with Mach, who was an influence on Schlick. Another influence on Schlick was Wittgenstein. Kripke (1982: Postscript) argues that Wittgenstein's no-bearer view of experience underlies a conceptual problem of other minds raised in the *Philosophical Investigations*.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

the inside falls short of capturing the intended non-solipsistic hypothesis that there are other experiencers. There seems to be a perfectly intelligible distinction between, on the one hand, a situation in which experiences that are causally dependent on two bodies still belong to one experiencer, and on the other hand, a situation in which experiences which are causally dependent on two bodies also belong to two experiencers. But without some notion of 'having' an experience besides causal ownership, these distinctions cannot be respected. Conversely, we seem equally able to make sense of the possibility that the experiences of two experiencers might in fact be caused by the condition of a single body. As we shall see in Section 4, this is a possibility that split-brain patients have been taken to actualize. It follows again that there must be a way of 'having' an experience which is distinct from causal ownership.

Another attempt to smoke out the conceptual inadequacy of the no-bearer theory is to ask, as many have of Hume's famous claim not to find a 'self' in the flow of perceptions: *who* is trying and failing to find a bearer of these experiences? *Whose* merely partial reflective perspective on all experiences is being expressed by the use of 'these experiences'? *Who* is finding that these experiences are causally dependent on a certain body? Again the natural answer is: whoever is undergoing or modified by the experiences: in short, their bearer.

The no-bearer theorist can respond. What becomes plausible under pressure it is that *some* further notion of ownership aside from dependence on a causal interface must be acknowledged. But it is not obvious that an owner in this further sense has to be something of which experiences are the undergoings, or modifications, or otherwise essentially inherent. For example, we could think of experiences as coming in multiple unified clusters or bundles. Experiences are bound together by direct causal interrelations with certain other experiences and not others. In response to the conceptual problem of solipsism, the no-bearer theorist can say that the intelligible distinction is just the distinction between the situation in which experiences that have two causal interfaces are elements of one bundle, and the situation in which experiences that have two causal interfaces are elements of two bundles. Conversely, as might occur in the split-brain case, we can make sense of two bundles being dependent on one causal interface. And in response to the challenge to explain who has the partial reflective perspective on 'these experiences', the no-bearer theorist can explain that one aspect of the causal interrelatedness of experiences within a bundle is the registration or indication of elements of the bundle by other elements of the bundle. Some experiences immediately give rise to and become the objects of a further thought 'these experiences'. That's all it comes to for some but not all experiences to be given to a reflective perspective.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

The appeal to bundles does not reintroduce bearers. The no-bearer theorist can reasonably say that neither *being included* in a certain interrelated plurality of experiences, nor, if we think of a bundle as literally one composite object, *being a proper part* of a certain bundle of experiences is part of the very nature of a particular experience. A given experience is only causally, and so contingently, related to any particular plurality or bundle of experiences. But a bearer, recall, would be something upon which an experience is ontologically dependent.

A variant approach develops a notion of owner as not a plurality or bundle of interrelated independent experiences, but as an abstract object whose identity is explained by an (equivalence) relation over independent experiences. Compare an 'orientation'. The orientation of one line is identical to the orientation of another line just in case the lines stand in the relation of being parallel. Or consider identity of the number as defined in neo-Fregean fashion: the number of *F*s is identical to the number of *G*s just in case the *F*s and the *G*s stand in the relation of equinumerosity. Similarly, the identity of a sort of 'owner', which we could call an experience's *abstract centre*, can be defined in terms of a relation of 'co-consciousness', where co-conscious experiences are phenomenologically unified with each other, composing a more inclusive conscious experience.[6] The abstract centre of one experience, then, is identical to the centre of a second experience just in case the experiences are co-conscious. Different experiences can 'present' the same abstract centre, as different lines can present the same orientation, and as different collections can present the same cardinal number. But an owner so conceived is no more than a kind of shadow of the co-consciousness relation between its experiences. The experiences are no more ontologically dependent on their abstract centre than a pair of apples are ontologically dependent on the number two. Abstract centres, again, are not bearers.[7]

To sum up our tour: we have encountered the locus of bodily sensation, causal interface, bundle, and abstract centre. Probably not even these exhaust ways for the no-bearer theorist to consistently acknowledge the 'ownership' of experience. Another relation is spatial containment. Hume believed that (most) perceptions in a bundle were literally nowhere in space. But a contemporary no-bearer theorist is

[6] Bayne (2010: Ch 1).

[7] This approach reconstructs aspects of Bayne's intriguing suggestion that 'the self' is 'virtual'. If the self Bayne has in mind is an abstract centre, then, as he claims, a self is not the sort of thing that *could* possibly have been found among the concrete objects in the world (2010: 293). In this respect a self is like a number or orientation, and unlike a dragon or a talking donkey. And Bayne's claim that it is a necessary truth that the experiences of a single self are co-conscious (2010: 281) will be correct, for the same reason that it is a necessary truth that lines which present the same orientation are parallel. The abstract object is defined by the relation.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

not committed to this form of dualism. They may think of experiences as spatially contained within brains and within animals for example.

2.3 Lonely Experiences

Now the Cheshire Cat's grin hovers back into view with a challenging objection. Since none of these owners surveyed is a bearer, it should be possible for an experience to exist all by itself, unowned, separated from any larger bundle, abstract centre, or causal interface with which it happens to be connected. But surely this is just as impossible as a lonely grin.

There are two kinds of responses to this objection. One response questions whether the no-bearer theory is really committed to the possibility of a lonely experience. Another response questions whether a lonely experience is really impossible.

The first response can draw upon Lichtenberg's riposte to Descartes's *cogito*: 'One should say *it is thinking*, just as one says, *it is lightning*. To say *cogito* is already too much as soon as one translates it as *I am thinking*.'[8] Notice that Lichtenberg here does not just question Descartes's presupposition that thoughts have bearers. His comparison with lightning gives a positive alternative to conceiving of thoughts as the undergoings or modifications of a bearer. A flash of lightning is an individual process (or perhaps a particularized property-instance or 'trope') which does not relate to any single object in the way that a grin relates to that which grins. Nevertheless, we might think that lightning is a physical phenomenon, in the sense that its occurrence necessitates certain physical realizing conditions. There must be the right atmospheric conditions, charged particles of a certain kind, a difference in electrical potential. It doesn't follow that there is an object upon which the flash of lightning is ontologically dependent. It depends by its nature on the obtaining of physical realizing conditions of a certain generic kind, but not on any particular bearer. In parallel, the no-bearer theorist can consistently reject the extreme Humean view that an experience could exist all by itself as the only thing in the universe. Its occurrence can be admitted to depend on physical realizing conditions of a certain generic kind.

The second response probes the basis for the claim of impossibility. Can't we conceive of a momentary 'phenomenal spark' in the void? Of course if it is all alone, then on the assumption that introspective reflection upon an experience requires some other element of a bundle to register the experience, then it cannot be introspected as it occurs. But first, one might think there is already empirical evidence of experience which escapes reflective access (Block 2011). Second,

[8] Lichtenberg (2012: 152). *Sudelbuch* K76.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

the insistence that an experience simply cannot occur isolated from a bundle might be unkindly likened to Berkeley's claim that an unperceived tree is inconceivable. Perhaps there is a kind of imagination of an experience 'from the inside' which involves imagining introspecting the experience, in the way that visualizing a tree arguably involves imagining seeing a tree from a certain point of view. In this sense of imagining experience from the inside, then, one will always fail to imagine a genuinely isolated experience, just as one will always fail to visualize an unseen tree. It doesn't follow that we cannot coherently conceive of an isolated experience in some other way.

But isn't it part of the very nature of an experience to contribute to 'what it is like to be the subject'? (Nagel 1974). Again, it might be responded, when we imagine an experience 'from the inside' we imagine it in relation to some larger system which is implicitly on the scene. It is plausibly part of the very nature of an experience that *if* it is in relation to such a larger system, then it makes a contribution to what it is like to be that system. So we cannot really imagine an experience except as contributing to what it is like to be something. It doesn't follow that an experience cannot exist without being an element of such a system.

Another natural objection to the possibility of a lonely experience can also be accused of confusing conditions for knowing an experience with conditions for the occurrence of an experience. Observation of the outward behaviour of a system is central to our ordinary way of knowing of the occurrence of an experience. It might be concluded then that the experience just is an undergoing or modification of the very system whose behaviour is manifest. But this does not follow. The no-bearer theorist can agree that we use behavioural criteria to establish what type of experience is now going on 'in' a system. But having identified the type of experience on this basis, the token experience can consistently be thought of as an independent object, merely causally connected to other experiences and to the body.[9]

To these defensive moves, we can add two further clarifications about the no-bearer theory.

First, it focuses on conscious sensory experience. It is consistent with the theory that other broadly psychological conditions are ontologically dependent on bearers. For example, someone's tendency to show off at dinner parties, their gloomy manner, their great gymnastic ability: we need not attempt to think of these phenomena as existing except as dispositions of particular human beings.

Second, it is consistent with the no-bearer theory that experiences are shadowed or correlated with experiential 'things' which *are* ontologically

[9] Ayer (1963) makes this point in response to Strawson's view of experiences as metaphysically dependent on persons.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

dependent upon particular objects and which therefore do have bearers in our sense. Take the *fact* that someone is some experiential way, for example, the fact that Tom smells a mouse. It is plausible that this experiential fact, like any similarly structured fact, has its subject essentially. It is a fact that by its nature involves Tom. But this doesn't entail that Tom's *experience* by its nature involves Tom. Why? Consider the analogous situation with meteorological facts, for example, the fact that Berlin is stormy. This fact by its nature involves Berlin. But the individual flashes of lightning whose occurrence in the vicinity of Berlin ground this meteorological fact are not themselves ontologically dependent upon Berlin. None are by their very nature related to Berlin; they just happen to occur there. Or take the fact that this pile of laundry contains socks. This fact is dependent upon the pile of laundry. But the individual socks whose inclusion in the pile grounds this fact are independent of the pile of laundry. The no-bearer theory may take a parallel view of experiences. Their occurrence in spatial, causal, or inclusion relations to a person grounds an experiential fact which is ontologically dependent upon that person. But the experiences themselves are independent, like the socks or flashes of lightning. They are not to be identified with facts.[10]

2.4 Pathologies of Self

We have seen how the determined no-bearer theorist can fend off various objections. But this is not yet to agree that the theory is well-motivated in the first place. The interesting motivation for the theory was the way in which experiences seem bearer-less to introspective reflection. There are two broad ways in which the critic could question this motivation.

First, they might question whether the 'subjective' or 'inner' perspective should be taken to be revelatory of the complete nature of experiences. Perhaps we are just looking in the wrong place for a bearer of experience and it is the 'objective' or 'outer' perspective on the metaphysics of the mind which makes possible the principled identification of bearers of experience. For example, from a physicalist perspective it may look plausible that experiences are modifications of the particular brains (or specific brain parts) which ground or realize them. From a functionalist or teleo-functionalist perspective it may look plausible that experiences are by their nature the undergoings of functional systems such as organisms. These positive theories and their connection to objective theories of the mental will be discussed in Sections 3 and 4.

[10] See (Taylor 2020) for the contrary view that experiences are a lot like structured facts or states of affairs, and for just this reason ontologically dependent upon their subjects.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

But a second line of critical questioning engages with the motivation on its own terms, staying within the 'subjective' or 'inner' perspective and asking whether experiences really do seem to be bearer-less to introspective reflection.

Disagreements about the way things seem are notoriously difficult to adjudicate. But important recent work reaches out to clinical evidence about abnormal cases to support the theory that experience normally involves a ubiquitous 'sense of self': a special phenomenology which, like the background hum of the fan, becomes salient only when switched off. Philosophers have attempted to interpret patients' unusual reports in a range of cases as intelligible responses to the absence of an ordinary sense of self. Cases include the following: schizophrenic delusions of alien thought insertion (Duncan 2019); depersonalization – including Cotard's syndrome, in which patients go as far as to deny that they exist (Billon 2015); and advanced meditative and classical psychedelic experiences of 'ego loss' (Milliere 2020).

This genre of argument is currently popular but it faces challenges. First, it needs to be explained why the reports should be taken to be rational responses to some unusual phenomenology rather than simply irrational responses to the usual phenomenology. Second, even if it is granted that there is an unusual phenomenology, it is not obvious that this reflects the absence of something which is present in the normal case: the alternative is that patients' reports are responses to something positive which is absent in the normal case: for example, in the case of schizophrenia one might posit a positive phenomenology of alien intrusion or disruption, rather than the absence of a supposed feeling of 'mine-ness' which is normally present. Psychedelic experience of liberation from 'ego' might be understood to be a positive sense of liberation from fixed patterns of experience and thought, where this is freedom from a kind of limitation or constraint of which one has no awareness whatsoever in the ordinary case.[11]

But grant these pathological cases are agreed to reveal that a sense of oneself is ordinarily present. It remains a long road to argue that this is the sense of a *bearer* of experience, something to which a particular experience is by its very nature related. Our survey of other forms of ownership suggests alternatives. What breaks down in some unusual cases (e.g. the 'boundless' consciousness of advanced meditation) might be the usual sense of the bounded thing into which one's sensations fall, a sense of what we earlier called the locus of bodily sensation. Or what might break down is the usual sense that there is a single object upon which one's experiences are especially dependent: the sense of a causal interface. Or what might break down is the usual sense that an

[11] See (Letheby 2021) for philosophical discussion of 'ego dissolution'.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

experience or thought is causally related to the rest of one's mind: what one loses is the sense of its inclusion in a larger bundle of causally interrelated experiences and thoughts.[12]

2.5 The Subject in the Metaphysics of Sensory Experience

A deeper way of questioning the motivation for a no-bearer theory focuses on the picture of sensory experience which typically motivates the theory in the first place. Here the debate about bearers of experience interacts with analytic philosophy of perception.

In his otherwise uncritical summaries of Wittgenstein's 1930s lectures, G.E. Moore took exception to Wittgenstein's continued endorsement of a no-bearer view of experience, the view which he had memorably supported in his *Tractatus Logico-Philosophicus* by saying that 'the eye of the visual field' is not anything which is *in* the visual field. In the lectures, Wittgenstein endorsed Lichtenberg's suggestion that thought and experience should be reported impersonally, as 'it thinks' in the style of 'it rains'. Here is Moore's comment about this:

> Like so many other philosophers, in talking of "visual sensations" he [Wittgenstein] seemed not to distinguish between "what I see" and "my seeing of it"; and he did not expressly discuss what appears to be a possibility, namely, that though no person enters into what I see, yet some "person" other than a physical body or a voice, may "enter into" my seeing of it. (Moore 1955: 14)

The distinction Moore is pressing is the distinction between the object of experience (what I see) and the experience of the object (my seeing of it). The empiricist picture of experience which initially motivated the no-bearer theory will be accused by Moore of failing actually to describe *experience*, instead describing only the presented array of *objects* of experience. The taste of coffee, the afterimages, the sound of an engine, the itchy feelings in my limbs: these are *what* I am aware of. It is true that these data are subject-less; at best certain given objects (body parts, one's own voice) emerge as nearby and generally present, and privileged for the prediction of the further succession of the array of given objects. It doesn't follow that the experiences are subject-less.

Moore had influentially argued in his earlier 'Refutation of Idealism' (Moore 1903) that there is indeed more to a sensory experience than its object. On his *relationalist* view, each sensory experience is a complex situation, in which something, which he called a 'consciousness' or 'subject', is related in a distinctive way to an object. Russell would soon popularize the term 'acquaintance' for this

[12] Campbell (1999) explains thought insertion as a failure to monitor the causal dependence of one's thought upon a background of other mental states.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

distinctive relation between subject and object. An experience is comparable in this respect to a marriage, or a collision. The event of Alice and Bob colliding by its nature has both Alice and Bob as constituents. Likewise, an experience by its nature has a certain subject and object related in a certain way.

Notice how on this relationalist view, the subject of an experience cannot be a 'bundle' which has the experience as one of its constituents, and nor can the subject be a body with a merely external causal influence on the occurrence of experiences: on the contrary, on the relationalist view each experience has its subject as one of its essential constituents. The view has the consequence that experiences have bearers.

Relationalist views of sensory experience have had a resurgence in recent philosophy of perception. Some of Moore's early followers held that the objects of sensory experience must be objects somehow more private than the ordinary mind-independent environmental objects which are the focus of our everyday thought and action. But recent versions of relationalism about sensory experience typically hold that experience relates a subject to mind-independent objects and qualities (Campbell 2002; Martin 2002; Brewer 2011; Soteriou 2013).

But how is relationalism supported? It is typically supported on the basis that it captures the way experience naïvely seems. If one is asked to reflect on what one's current visual experience is like, one will look at the objects and qualities in one's environment, things with which one is perceptually related. The way we naïvely reflect on an experience presupposes that it is a situation in which a certain sector of the world is presented to the subject, or to put it another way, in which the subject is acquainted with a certain sector of the world.

Furthermore, some relationalists have argued that although one finds oneself attending to the objects to which one is perceptually related, one can also attend to aspects of the *way* in which one is related to them. For example one can attend to the spatial boundaries of one's visual field, but these are not features of some object of which one is visually aware (Soteriou 2013). Introspectable structural or perspectival differences between the sense modalities are naturally thought of as differences in the way in which one is acquainted with objects rather than differences in the objects themselves. This is further introspective support for relationalism.

A competing view is that sensory experience is a non-relational modification of a subject. Against Moore's contention that the element common to the sensation of blue and the sensation of green is the distinctive relation of subject to object, his 'adverbialist' critics pressed the alternative view that the element common to the sensation of blue and the sensation of green was not a distinctive relation to an object but the monadic condition of being somehow intrinsically consciously modified (Ducasse 1942; Papineau 2021). The difference between the sensations is not a difference in their 'objects', but a difference in the specific determination

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

of the determinable monadic condition of sensing. Just as an episode of walking slowly is some entity's walking in a certain way, so a sensation of blue is some entity's sensing in a certain way: their sensing 'blue-ly'.

This monadic conception of experience still disagrees with a no-bearer theory. An experience is again held to be something more structured than simply the presence or occurrence of some subject-less data. It is a subject's being intrinsically modified in a certain way. Moreover, against the empiricist view that no bearer is given to introspective reflection, the adverbialist can argue that on their view introspection of experience just *is* awareness of the bearer of experience. Not because the subject is a bundle of experiences, or some extra element that pops up besides the experiences, but because the subject is that of which the experience is an intrinsic modification. For comparison: when one sees a table's being a certain shade of brown and shaped a certain way, one just is thereby aware of the table which is so modified. A table is not somehow visually hidden behind its intrinsic modifications. There are of course differences between paradigmatic visual awareness and introspection: a visible table is usually given as a bounded thing selectable from a larger field, whereas in introspecting experience the subject so modified is not thereby given as bounded against a larger field of such things. But even if for this reason we decline to call introspection 'perceptual' awareness of the subject, it would still seem on the adverbialist view to be awareness of the way an object is in itself and therefore plausibly awareness of that object (Chisholm 1969).

A relationalist view of sensory experience does not support the view that we are aware of the bearer of experience in the same way. If experiences are relational situations rather than intrinsic modifications, then introspective awareness of an experience is not awareness of the way something is in itself. Yet it is plausible that being aware of a thing requires some awareness of the way it is in itself. So on the relationalist view the subject is at most something implied by what one is introspectively aware of, especially by the structural or perspectival aspects of experience, without itself being something one is introspectively aware of. Comparison: when you see an arc of a circle you see a situation in which things are related to a central point without necessarily seeing that central point.

But the relational view of sensory experience may actually be more robust than the adverbialist view in the face of no-bearer opposition. A natural way for the no-bearer theorist to develop the theory that experiences are ontologically self-standing things is along 'trope theoretic' lines. According to a trope theorist, what we think of as the monadic situation of a certain ordinary object being, for example, red or dented, should be understood fundamentally as the occurrence of a primitive and unanalysable quality instance, *this red-ness*, *this*

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

dented-ness, rather than a complex situation in which a particular substance has a monadic property or universal. A familiar object like a table then is a sort of bundle of these independent quality instances which are jointly 'compresent'. The no-bearer theorist about experience can apply this general metaphysical view to experiences and the 'minds' they compose.[13]

The adverbialist view is that experience is a certain monadic situation, for example, the exemplification of the determinate property of sensing blue-ly. This is not the property of *being* blue; the view is not that the subject is literally blue. But insofar as a trope theoretic reconceptualization of monadic situations is generally available, there is no special obstacle to the no-bearer theorist reconceptualizing this situation as simply an unanalysable particular instance of sensing blue-ly. This is a bearer-less particular; a spark of phenomenal blue. But it is far more difficult to analyse a *relational* situation in a way which dispenses with its separate relata. It is hard to make any sense of self-standing instances of *picking-up* or *being-next-to*, or *colliding* without a commitment to a pair of things which are thereby related. It is not surprising then that trope theorists have either conceded (contrary to the usually atomistic spirit of the theory) that there are 'relational tropes' which do after all necessitate the existence of their relata,[14] or else have sought to eliminate relations altogether.[15]

This is not the place to adjudicate monadic and relational views of sensory experience. The aim here is just to suggest a way in which the question of whether and how experiences have a bearer relates to other issues in the philosophy of mind, and the philosophy of perception in particular. A relationalist view of experience quite strongly supports the existence of a bearer. A monadic view of experience is less resistant to a fundamental analysis according to which experiences are like other monadic situations, just the occurrence of subject-less quality instances.

As suggested, relationalism does not imply anything further about the nature of the subject-relatum of experience. If experience is wholly relational then introspection of experience arguably *could* not reveal what the subject is like in itself. So even if the naïve inner perspective supports the view that there is a bearer, an outer perspective may be required to determine what kind of thing it actually is.

2.6 Ourselves

The previous section has been concerned with the question of whether experiences have *bearers*: whether experiences are ontologically dependent upon

[13] For a trope interpretation of Hume's subject-less 'perceptions' see (Hakkarainen 2012).
[14] Maurin (2011) and Wieland & Betti (2008). [15] Campbell (1990) and Simons (2016).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

experiencers. How does this question relate to the question of 'self' with which we began, the question of what we are?

The questions are closely related on the traditional assumption that we are the bearers of our experiences if we are anything at all. If the bearers of our experiences are souls, then we are souls. If the bearers of our experiences are rational animals, then we are rational animals. But the traditional assumption has a converse implication: a no-bearer theory entails a 'no-self' theory: if our experiences – or perhaps better *these* experiences – have no bearer, then there is nothing that we are.

But a no-bearer theory may dispute the traditional assumption. With various other senses of ownership now elucidated, it is an option that the reference of 'we' – or for each of us individually, the reference of 'I' in self-representational thought or speech – might be an object or system which 'owns' those self-representations without being a bearer. Self-representations plausibly pick out the very system which in *some* sense 'has' them, but the self-referring system could be regarded as, for example, the bundle which contains the representations as parts – or perhaps the larger composite which is composed of the bundle and the causal interface with which it is intimately related: 'the mind plus the body'.[16]

There is no quick route from a no-bearer theory to a no-self theory. Is there any other route to the view that there is nothing that we are?

2.7 Functional Role and Reference

Self-representations – whether fully conceptual 'I' thoughts, or non-conceptual analogues – are representations with a certain distinctive functional role. They function to directly monitor and control the very system which owns them, the representing system itself. In harmony with this functional role is the principle that if self-representations function to monitor and control a certain system, then they refer to that system. Self-representations are in this sense about the representing system.

Anscombe (1975) denied this principle. She gave a pioneering description of what she called 'self-consciousness' or what we have called the functional role of 'I' thoughts: the ways in which such thoughts are characteristically sensitive to certain immediate informational inputs from the owning system and have immediate influence upon the behaviour of the system. Famously however she argued that 'I' does not refer to anything.

What makes sense of this surprising-sounding view is her background conception of referring as an activity requiring the sortal-guided identification of

[16] This latter is Parfit's (1984) 'reductionist' view of what a person is.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

a referent.[17] It is plausible that first-person thought does not require sortal-guided identification of a referent. First, it seems that one can competently use the first person while being agnostic about what sort of thing one is, about what 'sortal concept' one falls under. Indeed the very existence of radical philosophical disagreement about the sort of thing we are is suggestive of the sortal independence of self-representation. There is no similar disagreement about whether for example 'Kilimanjaro' refers to a material thing or an immaterial thing. Here the referring expression does have a closely associated sortal concept (mountain).

Second, there seem to be 'identification-free' uses of the first person for which one doesn't need to look around and identify the referent for one's use of the first person. Many philosophers agree that this identification-freedom is evidenced by the impossibility of a certain kind of error. For example, when one comes to think on the normal basis, 'I am in pain' it does not seem that one could be right that *someone* is in pain but go wrong solely in identifying *who* is in pain. But one would predict such partial error to be possible if the basis for one's judgement involved an identification as a step: that person is in pain; I = that person; therefore I am in pain. One should be able to get the first step right but get the second identification step wrong.[18]

However, these peculiarities of self-representation are not really a solid basis for arguing that there is nothing that we are. We can distinguish Anscombe's *use-theoretic* notion of a referring expression from a more common *semantic* notion. Even if it is true that the use of 'I' differs from the use of paradigm referring expressions in not being guided by sortal-identification, it could still be the case that the expression refers in the sense that it makes the truth of one's statements containing the expression turn systematically on the condition of a single object.[19] If this object has a certain basic nature, then that is one's basic nature. There will be a truth about what one is.

2.8 Mismatch

Anscombe's position is that the philosopher who construes 'I' as referring expression is mistaken. She does not attribute any illusion to the ordinary thinker. But more recent writing suggests another way of denying the principle that if self-representation functions to monitor and control a certain system, then it refers to that system. This *mismatch argument* attributes illusion to the ordinary thinker.

[17] See (Haddock 2019) for usefully sympathetic interpretation of Anscombe.
[18] Shoemaker's (1968) pioneering reading of Wittgenstein's *Blue Book* initiated intense discussion of this phenomenon of 'immunity to error through misidentification'.
[19] As Anscombe herself appears to concede (1981: 32–33).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Here is an analogy to illustrate the idea. Consider the following way of controlling a guided missile on to its target. The military have found that an effective system is to dupe children into playing an online computer game in which they are rewarded for flying a cute cartoon bird back to its nest. The system is set up so that monitoring and control of the bird's flight to the nest works to monitor and control the missile's trajectory to target.

This is an example where a certain representation (the child's 'that bird') in fact has the function of monitoring and controlling a certain concrete object (the missile). Yet in this example it is to say the least not obvious that the representation refers to the concrete object. When the child points at the cartoon bird and says 'that bird' are they expressing thoughts about the missile? When they think 'that bird has yellow feathers' does the truth of what they say turn on whether the *missile* has yellow feathers? It would seem to be a stretch. The information which attaches to the child's 'that bird' representation concerns yellowness, feathers, beak, relation to a nest and other avian circumstances, none of which fits a missile at all. The correlation of apparent flight structure and actual trajectory structure seems insufficient all by itself for the child to be picking out the real missile in thought. So plausibly 'that bird' does not refer to anything – at least not to any real concrete object with a nature which goes beyond the graphical representations of it.

Some writers have taken an analogous view of self-representation. The idea is that although self-representation functions to monitor and control the containing system, it fails sufficiently to match or fit the system to count as referring to the system. So there is nothing that you are: at best you are a kind of 'virtual' or 'fictional' character, like the cartoon bird, defined by representations that a system uses to monitor and control a certain concrete object. The representation of a self is a 'user illusion' which enables a system to monitor and control itself.[20]

Illustrations of failure of match in the literature include narrative confabulation about one's past, character, and agency (Dennett 1992); the sense that one is located at a place somewhere behind one's eyes (Hood 2012); and the representation of oneself as belonging to a special metaphysical category, such as bearer of experience, indivisible simple entity, or bare particular (Johnston 2012). In each case, it is arguable that there is no real object that matches the self-representations in the relevant respect.

A supposed mismatch between representation and reality can be challenged in either of two ways. First, do we really represent ourselves that way? Second, is there really no object that is that way?

[20] For views in this spirit, see (Dennett 1992; Metzinger 2004; Ismael & Pollock 2006).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Here is an example of the first way of challenging a supposed mismatch: consider the argument that we represent ourselves as an indivisible metaphysical simple on the evidence that self-representation doesn't reveal any of the complex hierarchical sub-personal processing which underlies our self-representation. All of this complex processing in the end serves up one simple convenient 'I'.[21] On the face of it, this argument would be a non-sequitur: it moves from the (let it be granted empirically supported) premise that self-*representation* is simple to the conclusion that self-representation represents its object *as* simple.

Here is another example of the first challenge: it has been argued that we represent ourselves as exotic things which are distinct from any ordinary concrete object on the grounds of the psychological evidence that we can imagine ourselves 'from the inside' being anybody, perhaps anything with a point of view. I can imagine myself as Spinoza penning a letter, as Stephen Hawking solving equations[22], or as a butterfly flitting from flower to flower.[23] But no ordinary concrete thing has such a protean, mutable, nature that these are all real possibilities for it. So this reveals that I must think of myself as a metaphysically exotic 'bare particular' which can float free from any ordinary concrete thing.

But this too seems to be a non-sequitur, as the comparison with the use of another perspectival term in imagination brings out. I can vividly imagine that it's 2030 now, I can imagine that it's the Stone Age now, I can imagine that it's 1999 now. Quite obviously there is no real time which is at once possibly 2030 and possibly the Stone Age and possibly 1999. So is this evidence that I think of 'now' as picking out some exotic object which is distinct from any ordinary time? No. We all understand that 'now' usually picks out whichever ordinary time is present. But in describing the content of imagination it can shift to pick out the time at which imagined events are present. The parallel and reasonable view of 'I' is that it ordinarily picks out an ordinary concrete person, but in describing the content of imagination it can be used instead to pick out whatever in the imagined situation occupies the subjective view on the imagined events.[24]

Here is another example of shift of the reference of 'I' in counterfactual thinking. A mother tells her children not to open the door to anyone while she is out. But as she returns, they open the door to her. 'You shouldn't have opened the door!' she chides, 'I could have been the wolf!'.[25] Does this speech show that the mother tacitly conceives of herself as an exotic werewolf-like entity

[21] Metzinger (2004) and Letheby & Gerrans (2017).
[22] These examples are used by Johnston (2012: 197) and Garfield (2022) respectively in service of mismatch arguments for 'no self' theories.
[23] This example is from the fourth century B.C. Chinese philosopher Zhuangzi (Watson 1968: 49).
[24] See (Velleman 1996) for discussion. [25] This example is taken from Heim (2004).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

who could have transformed from human to wolf before returning home? Obviously not. There is a description, which the mother, the reference of 'I', saliently fits, 'the thing in the doorway', and the counterfactual speculation 'I might have been a wolf' is not about the actual reference of 'I' but shifts to the thing which counterfactually fits the associated description. What is being imagined is not that the mother herself is a wolf but that the thing in the doorway is a wolf. Similarly, one might argue that when imagining from the inside 'I might have been Spinoza' one is imagining Spinoza to be the thing which fits a condition one saliently fits. The condition here is not that of occupying a doorway but the condition of occupying the subjective point of view. One imagines from the inside a situation in which the thing which is gazing out from a certain perspective is Spinoza.

Here briefly is an example of the second way of challenging a supposed mismatch. Suppose it is agreed that we represent ourselves as the bearers of our experiences. It could be argued that this results in no mismatch because our experiencers really *do* have bearers, for example on the basis of the recently reviewed theory that a sensory experience is by its nature a relational episode in which a certain subject is related to certain objects.

2.9 Reference Determination

But some degree of mismatch is beyond question. For example, nothing quite lives up to the autobiographical tales we spin about ourselves. And perhaps we really are inclined in certain conditions to judge that we are at a point somewhere behind our eyes.

But it would not follow that self-representation picks out no real object. For it is possible to represent a real object, and also misrepresent it in some ways.[26] To illustrate this rather obvious point: perhaps no real human being lives up to the over-idealized way in which I think of my mother. No human being is really that kind, wise, and lovely. It doesn't follow that my mother does not exist, or is a fictional character. She is a real concrete human being, about whom I have a distorted conception.

Still, one might be puzzled. If my mother doesn't exactly fit my mother-representations then how does she nevertheless get to be the object of my mother-representations? Here we tread into the deep waters of reference-determination theory in general. But here is one promising general approach. According to a principle of *epistemic charity*, the right overall assignment of reference to representations is the assignment that best interprets the representer

[26] See (McClelland 2017) for development of this point in criticism of mismatch arguments about the self.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

as possessing knowledge of the things assigned. To illustrate: among the various relations in which I stand to my mother are epistemic relations. I have seen her, I remember her, I reason about her; chains of verbal and written testimony reach back to her. In virtue of these relations, the assignment of my mother to my mother-representations will ascribe to me some error, but also a lot of knowledge of what she is like and what she has done. And no other assignment will do better in the ascription of knowledge. So she is the reference.[27]

On this approach the right assignment of reference need not be a perfect fit. Nor is perfect fit, or even best fit, sufficient for reference. Suppose that there is someone on the other side of the universe who not only looks and acts like my mother but is closer than my mother to being as kind, wise, and lovely as I represent my mother to be. Intuitively my mother-representations are still not about this 'twin-mother'. The epistemic charity theory explains why. I am not in any kind of epistemic contact with this remote individual. I have never seen her or heard about her. So the assignment of twin-mother to my mother-representations would ascribe to me a lot of luckily true belief, but no knowledge whatsoever.

Another plausible consequence of the epistemic charity approach is that it permits even quite radical 'sortal error' about the referent as long as other compensating knowledge is ascribed: faced with a quite convincing artificial plastic plant, I can pick it out with 'that plant' and think about it, despite mistaking an artefact for a living organism. The assignment of the artefact to 'that plant' still interprets me as possessing plenty of knowledge of its colour, shape, and location for example.[28]

Does the epistemic charity approach explain the intuitive verdict about the cartoon bird and the missile? Yes. The child using 'that bird' knows nothing of the missile which happens to be causally correlated with their experiences in a certain minimal respect. One can envisage an operator who knows about the trajectory correlation and on that basis comes to know by description 'the missile which is correlated with these experiences in such a way is turning left . . . ' etc. But the child has neither direct cognitive contact nor indirect knowledge by description of the missile. In this case neither the missile nor any

[27] See (Williamson 2007: ch 8) for defence of a 'knowledge-maximization' approach to content determination. Recanati's (2012) view is that the reference of 'mental files' is fixed what he calls 'epistemically rewarding relations' – in effect ways of coming to know about the referents. Dickie's theory (2015) that the object of a body of beliefs is the object that renders the beliefs 'justified' and 'non-luckily true' studiously avoids appeal to knowledge but delivers similar results. For a congenial discussion of epistemic factors in fixing *de se* reference in particular, see (Morgan 2019).

[28] For more on reference without correct sortal classification see (Madden 2019).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

other concrete thing is a knowledge-ascribing assignment to 'that bird'. It is better treated like a fictional name.[29]

The self-representation case is arguably quite unlike this case and much more like the mother-representation case. There is a larger system which has a range of ways of knowing immediately about that very system. Visual self-location, autobiographical memory, proprioception, the testimonial uptake of others' 'you'-assertions: these are all what might be called reflexive channels of knowledge. If the larger system is assigned as the referent of self-representations, then the system is ascribed a lot of knowledge of itself, along with a certain amount of error. What is this larger system? The obvious answer is that it is a human being, a living organism of a certain kind. Reflexive monitoring, of internal homeostatic state and overall state of motion for example, is common to all animals, including human beings.[30]

Note that a channel which generally gives knowledge of a certain concrete object can be forced into error or 'mismatch' when abnormal precision is demanded. Here is an example: the icon on the dashboard GPS map represents the car's location. This is a charitable assignment of reference because the system is a basis for knowledge of where the car is on the map, at least within the range of precision of ordinary use. Suppose that the sensors actually triangulate to a particular point over the engine. Now suppose that the driver is prompted to use the GPS map unusually highly zoomed-in, to answer the quasi-philosophical question 'where *exactly* is the car?'. Peering at the icon on the screen and shuffling the car back and forth the driver gets the point over the engine to line up with a certain precise coordinate point. 'OK the car is just *there*: not one metre forwards, not one metre backwards'. But no concrete object is exactly there. So does this reveal that the icon represents a mere fiction, a kind of exotic point-like inner-car within a car? No. The concrete car remains overall a perfectly charitable assignment of reference to the GPS icon, despite the possibility of this kind of peripheral error when it is artificially pushed outside of its normal bounds.

Peripheral mismatch in visual self-representation is similarly predictable. We normally use vision to locate ourselves reliably with respect to moderately distant landmarks. I'm nearly at the top of the hill. I'm between two chairs. We do not normally use vision to answer the question: where am I with respect to two fingers held very close to my eyes? Or: where I am with respect to my closed eyelids? But if prompted by experimenters to do so, or if in an odd

[29] We need not address the question here whether fictional names refer to abstract objects or refer to nothing at all.

[30] See (Jékely et al. 2021) for a recent discussion of the significance of self-monitoring of motion and bodily configuration ('reafferent sensing') in the evolution of animal life.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

philosophical mood one prompts oneself, then, given the triangulating role of the eyes in visual self-location one might be inclined to say that one is at some point behind one's eyes. Still the larger human being remains the perfectly epistemically charitable assignment to self-representations, despite the fact that on this assignment of reference visual self-location is counted as error-prone when forced outside of its normal operating range of precision.

To sum up, the 'no-self' view that our self-representations refer to no real object faces an uphill struggle to leverage support from the sciences of the mind. Even if there really is a mismatch between our self-representations and any real concrete candidate assignment of reference, the plausible view remains that the charitable assignment of reference to our self-representations is a human being, a certain kind of self-monitoring and self-controlling animal.

A last hope for a mismatch argument is the thesis that there are actually no such things as animals. Then our self-representations could after all have a kind of hallucinatory nature. The compositional nihilist view that there are no composite objects of *any* sort would be a principled basis for this claim. Compositional nihilism is a non-arbitrary answer to the question of when things compose a further thing (answer: never) and it has ontological economy in its favour. It is a serious view.[31] But here we will set aside this sweepingly revisionary picture, and explore some more focused challenges to the hypothesis which has now emerged as plausible. This is the hypothesis that we, the objects of our self-representations, are animals of a certain kind. Those with compositional nihilist scruples are free to think of the entirety of what follows as prefixed by the fictionalist operator 'According to the fiction that composition occurs . . . '.

3 Persistence

Something persists just in case it continues to exist. Persistence is central to the contemporary debate about the 'animalist' view that we are animals of a certain kind. In short, there are some cases of persistence which seem to support animalism over its rivals; and there are some cases which generate intuitive verdicts widely thought to be problematic for animalism.

In this section, we will review some of this debate and ask whether the latter cases are quite so problematic for animalism. Recently a version of animalism has been proposed which can accommodate intuitions on both sides. However, there are some special objections to such a version of animalism which deserve to be aired. Such a version has been accused of being unprincipled, and of having implausible metaphysical consequences due to its 'best candidate' structure. The section will end with some responses to these objections.

[31] Dorr & Rosen (2002).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

3.1 Animalism and Brain Transplants

Animalism is not a claim about persistence. It's a claim about what we are. It says that we are certain animals. This can be understood as a plural identity claim: there is a certain plurality of animals, such that every one of us is among them, and every one of them is among us. We are they.

Why believe this claim? As we will see below, a sophisticated 'thinking animal' argument has been endorsed by some authors. However, more basic grounds can and have been given. It is noted that there is an apparent coincidence between ourselves and certain animals in all ordinary respects. Wherever we go, in whatever condition, there go some animals. The simplest explanation for this coincidence is that we are those animals.[32]

However, this way of describing the situation rather understates the motivation for the theory, suggesting that it is like the discovery that the Morning Star is the Evening Star, or, to use a plural comparison, the discovery that the men dressed as window cleaners in the morning are the men who robbed the bank at night. In such cases, we first have two ways of thinking of objects and then we discover that these two ways of thinking in fact home in on the same objects. But it's quite contrived to suppose that we have a conception of 'these animals' or 'our animals' on the one hand, and a conception of ourselves on the other, such that we then find evidence that they are in fact the same things. We don't distinguish ourselves from animals in the first place. Rather we seem to have learned, *a posteriori*, that we, the objects of our self-representations, are animals, things of a broadly similar nature to other denizens of the biological world. So the question is not: what reason is there to merge our conception of ourselves with our conception of 'these animals'. The question is for the critic of animalism: why split off an artificial conception of 'these animals' from our self-representations in the first place?

In this respect, animalism importantly contrasts with the claim that we are 'our bodies'. There is a scientific sense of 'body' which means cohesive material unit (as in: the electrodynamics of moving bodies). But only things with a psychology can sensibly be said to *have* bodies. A rock is a body in the scientific sense, but it doesn't *have* a body. In the sense in which one has a body, one's body is the complement of one's mind: it is roughly the non-psychological side of something which has or normally has a psychology. And we do ordinarily conceptually distinguish ourselves from our bodies in this sense. It is not a default starting point that we are identical to our bodies. Far from it. The claim that we are identical to our bodies actually sounds *prima facie* paradoxical. It would take a great deal of explanation to defend it.

[32] Snowdon (2014: Ch 4) and Bailey (2015).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Still, sometimes the occasion does arise to split up what was initially a single conception. Suppose I am holding a pen. But then strangely I find that when I physically wiggle what I feel in my hand, the pen I see appears not to move. Could it be that a prism is making a distant pen appear to be at the location of the pen in my hand? I need to separate a conception of 'this pen' from a conception of 'that pen' in order to frame the identity question which has arisen thanks to the apparent instantiation of the incompatible properties of moving and not moving. Is this pen actually one and the same as that pen?

The apparent instantiation of incompatible properties is at the heart of the two most important reasons philosophers have given for splitting a conception of 'these animals' or 'our animals' away from our self-representations. Unsurprisingly these appeal to exotic situations; if the appearance of incompatible properties was ordinary then we might already talk separately of 'our animals' or 'these animals'. First, unusual cases of disunity have been used to argue that among us are things which are discernible from animals. Disunity is a topic of Section 4. Second, counterfactual 'Lockean transfer' cases have been used to argue that we have persistence conditions that animals do not: it is argued that there are circumstances in which one of us would continue in a certain state, yet no animal would continue in that state. In this section, we will focus on the transfer case which has been most influential in twentieth and early twenty-first century debates (Shoemaker 1963).

Suppose that your cerebrum, or whatever is the part of your brain whose operations physically realize your mental phenomena, is moved intact and functioning from your skull into the skull of an animal whose cerebrum has recently been destroyed. The animal which was the best candidate to be you before the operation is apparently left behind by this procedure: although left in a vegetative state without a cerebrum, it continues to breathe, metabolize, and so on. On the other hand, at the location of the recipient animal now is someone who answers to your name, appears to have memories of your childhood, knows your passwords, has exactly the same taste in music, and so on.

Most people who encounter this case find it extremely plausible to say that you would 'go with' the cerebrum, perhaps in some combination of the following senses: you would continue to be the bearer of the mental processes physically realized by operations of the cerebrum. You would be the subject of sensory experience subserved by the cerebrum. The first-person speech controlled by the cerebrum would still to refer to you. You would be at least partly located at the location of the cerebrum.

If this kind of transfer between animals is a possibility for each of us, while it is not a possibility for any animal, then we must be distinct from animals.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

We will next sketch some alternative views of what we are, but then raise some problems for these rivals.

3.2 Alternative Views and Their Problems

Suppose that we accept that the possibility of cerebrum transplantation demonstrates that we are not animals. What else could we be?

One answer is that we are souls as conceived in the Platonic and Cartesian traditions. We are immaterial things, which, by cerebrum transplantation, could come to be attached to different animals at different times. We won't explore this option beyond noting that it is of mainly historical interest, and faces the well-known challenge of explaining how we could possibly causally interact with the material world. The variant view that we are a union or compound of soul and material body, and not just a soul, only relocates the problem of causal interaction to somewhere within our own boundaries.

Are there alternatives to animalism which are materialist? The cerebrum transplant case might suggest the view that we are parts of animals, the parts which would get moved between animals in this case. We are brains or cerebra.[33]

While this view is consonant with a certain kind of 'neuromania' in popular intellectual culture, it is in fact a very strange view. It fits with the intuition that we could be transferred but it fails to fit with far more ordinary aspects of our self-representations. We represent ourselves as larger than a few inches in height, and as visible to one another and not hidden behind our faces. And even if the brain theory fits with the transfer intuition it has other implications about our persistence which are questionable. Suppose that one's brain is preserved in formaldehyde in a museum after one's death. If you are identical to your brain then this is a situation in which you continue to exist, perhaps for many years after your own death. This is not a particularly appealing implication of the view. That said, we will encounter some principled reasons for favouring something like the brain view in Section 4.

Probably still the most popular materialist alternative to animalism is what can be called neo-Lockean coincidentalism.[34] According to this view, we are neither identical to animals, nor are we proper parts of animals. We stand to them in the different but still intimate relation of coincidence. This is easiest to explain with an example. Suppose that a sculpture has been made by shaping a piece of clay. It can be argued that the sculpture and the piece of clay are numerically distinct objects. This is because they are discernible in various respects. For example, the piece of clay is older than the sculpture. The

[33] Searle (1983: 230); McMahan (2002: 88–94); Tye (2003: 142); and Parfit (2012).
[34] Baker (2000) and Shoemaker (2008).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

sculpture, but not the piece of clay, could be well-made, or classical in style. The piece of clay would continue to exist if the clay were to be flattened out into a sheet but the sculpture would not continue to exist in those circumstances. It would be destroyed. Conversely, suppose that the sculptor maintains the sculpture over a long period by gradually replacing small tarnished bits of clay and incinerating them. Eventually, it seems plausible to say that the original piece of clay no longer exists. But the sculpture still exists. It now coincides with a distinct piece of clay. In sum, the sculpture could go on without the original piece of clay, and the original piece of clay could go on without the sculpture

Although the sculpture is discernible from any piece of clay, it is no less a material thing. At every time at which it exists, it has exactly the same small material parts as a piece of clay. Neo-Lockean coincidentalism says that we stand to animals in the way that the sculpture stands to the piece of clay. We are material things at the same location as animals, sharing their matter, but discernible on the basis of a difference in persistence conditions. Animals have non-psychological biological persistence conditions. We have psychological persistence conditions: we continue to exist if and only if our psychological lives continue. Severe cerebral damage which extinguishes one's psychology is comparable to the squashing down of the piece of clay: there is no longer one of us coincident with the damaged animal. Cerebrum transplantation is comparable to a given sculpture first coinciding with one piece of clay and then coinciding with another.

Neo-Lockean coincidentalism faces a range of objections.

First one might oppose the very idea that distinct material things could coincide.[35] But such an opponent will have to explain away more than just the sculpture case. The natural world presents other apparent examples of material coincidence. We can distinguish a plant from its main stem. The plant has leaves and roots which the stem does not. It is larger. But now suppose that we trim away all the leaves and roots attached to the stem. It is plausible to say that the plant has not been destroyed as a result of this process but has simply grown smaller. Nor has the stem been destroyed by the process. So plant and stem now coincide. This period of coincidence may be short-lived; with water and nourishment, the plant will once again grow roots and leaves and become larger than its main stem. But for a while, two things exactly share the same matter.

Still, some may find it puzzling that two things exactly similar in their tiny material parts and the internal arrangement of those parts could possibly differ in any further way. What could ground or explain these supposed differences?

[35] van Inwagen (1990) and Ayers (1991).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

The shrunken plant and the stem are microphysically indiscernible for a period. So why does the plant then get larger while the stem stays where it is?[36]

In response, coincidentalists should object to the presupposition that an object's properties must be wholly explicable by its microphysical profile. Some of its properties are explained by its sort. For example, one object can grow leaves and roots because it is a plant. A plant just is the sort of thing which can change in that way. The other object cannot grow leaves because it is a stem. A stem just is the sort of thing which remains stem-shaped even when supporting other parts of the plant.

The objector might follow up: but now what explains why one object is a plant and the other is a stem? Here the coincidentalist may reply that an object's sort is simply part of what it is to be that particular object. It is a brute fact that it is of the sort that it is. Take the fact that a given particle is an electron. We do not expect this to have any deeper explanation. Of course, we might wonder why an electron has shown up and not a particle of some other kind. But it does not seem sensible to wonder of a particular electron why *it* is an electron. There's no evident reason for taking a different attitude to things of macroscopic sorts. The necessary conditions for the existence of a plant at a certain location may be more complex than the necessary conditions for the existence of an electron, but it is no less misguided to ask of this particular plant, why is *it* a plant?

The coincidentalist picture, then, is that coincident objects share some but not all of their properties. Common internal material profile determines that some properties are shared. Sortal differences determine that some properties are not shared; they are proprietary to certain sorts. For example, microphysically indiscernible plants and stems weigh the same and are the same overall shape. But only the plant can flourish, just as only the sculpture can be classical in style.

An important objection to neo-Lockean coincidentalism grants all this, but focuses in on psychological properties, and asks whether these are proprietary to certain sorts of things. Since kind determines persistence conditions, it follows that if it is right that we differ in our persistence conditions from animals, then we must belong to some other kind. Following Locke, it is common to call this putative kind *person*. Since we know that we have psychological properties, we know that persons can have psychological properties. But are psychological properties proprietary to persons? Or are they shared with the animals with which persons coincide? It would sound wrong to describe a stem as flourishing. It would sound wrong to describe a piece of clay as classical in style. It would also sound wrong to describe a lump of flesh, or one's body, as

[36] This kind of puzzle about coincident objects is generally known as the 'grounding problem'. See (Bennett 2004).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

perceiving and acting. Unfortunately for the neo-Lockean it does not sound at all wrong to describe an animal as perceiving and acting. Animals are quite ordinarily ascribed psychological properties on the basis of their behaviour and their presumed normal internal functioning. Experiencing would seem not to be proprietary to the non-animal persons that we are supposed to be. But if this appearance is taken at face value then neo-Lockean coincidentalists are committed to the presence of two experiencing things at one's location. Not only are there two things with exactly similar weight or shape. There are two things with exactly similar experiences.

This is known as the 'thinking animal' problem for neo-Lockean coincidentalists. Since it would appear to arise *mutatis mutandis* for any theory which distinguishes us from the animals in our vicinity, it can be presented as an argument for the view that we are these animals (Olson 2003).

One response is to deny that there is a real problem here. Compare: it would sound strange to be told that there are two objects each weighing 150lbs on the scales when the scales only read 150lbs. But once it is explained that we are talking about two objects which share their matter then at least some of the sense of strangeness dissipates. Likewise, it may sound strange to be told that there are two experiencers where you thought that there was only one. But once it is explained that these are coincident thinkers which therefore contain the same 'phenomenal sparks' (picture some of the shared matter as 'glowing') then perhaps some of the strangeness goes away. There may be two experiencers but at least it doesn't follow that there are two sets of experiences.

This response is most naturally paired with a no-bearer theory of experience in the sense of Section 2. It is not in the nature of a given experience to be related to a particular bearer. There are simply bundles of experiences, and these can be spatially contained within multiple objects at a single time.

However the fact it is experience rather than weight which is shared creates a special epistemic problem for the neo-Lockean. Grant that in your vicinity there is something with psychological persistence conditions, coincident with the animal which has non-psychological persistence conditions. The neo-Lockean claims that you are the person. But how could you possibly know which one you are? If the animal also thinks and experiences then you cannot use the fact that you are thinking to discriminate yourself from the animal. Psychological persistence conditions seem to be the only difference. But when, on the basis of reflecting upon the cerebrum transplant case say, you self-ascribe psychological persistence conditions, two thinkers self-ascribe psychological persistence conditions. The person is correct to do so. But the animal is not. So even if you are lucky enough to be the one who is correct, you have arrived at the truth on a basis that could just as easily have led to error. Knowledge requires

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

a safer basis than this. So, even if we are neo-Lockean persons, we cannot know that we are.

Some neo-Lockean coincidentalists have denied the apparently obvious truth that animals experience. They have claimed that psychological properties are proprietary to things with psychological persistence conditions. An argument for this conclusion has been developed with some sophistication by Shoemaker.[37] He draws on a functionalist theory of the mind according to which the possession of mental states involves the disposition to cause successor psychological states in the same subject. In a cerebrum transplant case, the earlier states of the animal fail to produce the appropriate successor states in the animal. Therefore, he concludes, animals are not subjects of mental states.

The most serious difficulty with this approach is to extract from a plausible form of functionalism anything stronger than the requirement that a mental state is one which *normally* or in the *right conditions* causes certain successor states in the same subject. The fact that an animal can fail to exhibit the appropriate successor states because, extraordinarily, it had its cerebrum removed hardly shows that it does not have states which would ordinarily cause the animal to go into the appropriate successor states at the later time. In Section 4 however, we will present a teleo-functionalist cousin of Shoemaker's approach which is quite promising as a solution to another problem of 'too many thinkers'.

Another neo-Lockean coincidentalist response accepts that animals think but uses a claim about first-person thought to address the special epistemic problem. It is claimed that a first-person thought does not simply refer to its thinker (on the present view there is no single thinker). Rather it refers to the thinker with psychological persistence conditions, the person. So, when both the animal and the non-animal use the first person, both refer to the thing with psychological persistence conditions. So the presence of the thinking animal does not undermine the safety of the claim that I am the person not the animal. Whichever thing is regarded as making the claim, it is a true claim, about the person.[38]

The challenge for this approach is to provide grounds for the theory that first-person thought is restricted to referring to the thing with psychological persistence conditions. If self-representations are monitoring and controlling two coincident-containing systems then why should they select just one of them? To the extent that Anscombe's contention that self-representation is sortal-free is plausible, it is implausible to suppose that it is built into self-representation to pick out just one of these two sorts of things.

But could the principle of epistemic charity introduced in Section 2 make the selection? Insofar as we are inclined to believe that we would go with the

[37] Shoemaker (1984, 1999, 2004). [38] Noonan (1998, 2010).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

cerebrum, wouldn't the thing with psychological persistence conditions be a better fit for our self-representations than the animal, and therefore the right assignment of reference?

This approach does not look promising. First, it is unclear whether the assignment would ascribe more knowledge. How exactly are we supposed to know that we are the ones who would go with the cerebrum? More importantly, now is the time to observe that the thinking animal problem is not the only problem for neo-Lockean coincidentalism. Even if it fits with the cerebrum transplant intuition, the claim that the continuation of a psychological life is necessary for us to remain in existence from one moment to the next is inconsistent with other plausible-sounding claims about our persistence. For example, if a loved one has lapsed into a persistent vegetative state, it is unlikely we would say that they have literally ceased to exist and only some object distinct from them lies breathing and snoring in the hospital bed. We might crave more time by their side before the life support is switched off. Each of us could, it seems, continue to exist without any psychology (although of course not for long without a great deal of assistance). It also seems that we persist through the earliest period of our life without any psychology. Each of us was once a twelve-week-old foetus, yet to develop the brain structures usually supposed necessary for sentience. Or so it seems.[39]

It is a cost of neo-Lockean coincidentalism that it has to deny these judgements. And given that we are inclined to believe that we could persist in non-psychological conditions, the epistemic charity approach to the determination of reference of our self-representations does not clearly favour the Lockean person over the animal. The animal better fits our judgements about our actual and possible persistence in foetal and vegetative situations. Once two coincident thinkers have been admitted, each of which fits self-representations in some respects but not others, indeterminacy of reference looks more likely. So the coincidentalist picture becomes in effect another route to a kind of 'no-self' theory. Our self-representations fail to select anything in particular.[40]

3.3 Denying the Brain Transplant Intuition

So we have surveyed some of the difficulties generated by the denial that we are animals. This suggests that we should scrutinize rather more closely the cerebrum transplant possibility which is commonly supposed to be inconsistent with animalism. If we could after all maintain that we are animals in the face of this

[39] Olson (1997).

[40] Sider (2001) develops an argument from a plurality of candidates which each fit in some respects but not others to indeterminacy of reference of personal vocabulary.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

intuition, then we could avoid these difficulties. There would be no need to argue that animals can't think, or that they can't self-refer. There would be no difficulty in supposing that we were once foetuses and that we could go into a persistent vegetative state. These seem to be possibilities for animals.

Animalists have typically challenged the intuition that we would go with the cerebrum. How? One way to go is to use the plausibility of animalism, enhanced in the light of the difficulties faced by the alternatives, as an argument against the truth of the intuition. Since animalism is true, it must be false. Our intuition here is just another example of a peripheral mismatch between our self-representations and reality. This may in the end be a reasonable response, but it is difficult to evaluate for two reasons. First, it is unclear whether the difficulties with the alternatives are all that severe. For example, some may find the consequence of the brain theory that we are only a few inches tall to be less counterintuitive that the denial of the transplant intuition. Second, this way of turning the anti-animalist 'modus ponens' into an anti-transplant 'modus tollens' relies on the same middle premise: no animal would go with the cerebrum. This response is only as strong as that premise. We have yet to evaluate that premise.

Another move is to weaken the intuition by presenting the same case in a different way. Bernard Williams once challenged the 'brain state transfer' intuition in this way. Although generally considered weaker than the cerebrum-transplant intuition, we have some inclination to judge that you would 'go with your psychology' in an imaginary scenario in which your psychology is recorded, your brain is erased, and the information transferred to another previously erased brain; we have some inclination to judge that the person at the end of this process, who answers to your name, and who seems to know your deepest secrets, is you. Williams changed the presentation of the case from the information-donor's point of view to the information-recipient's point of view.[41] If told that you will be tortured after being implanted with an entirely new set of memories and character-traits, then you will still be likely to fear the future pain. This alternative presentation of the case, then, makes it quite intuitive that you will not be replaced by somebody else, but will continue to exist, able to feel later pain, despite the otherwise massive psychological discontinuity induced by the interference.

The intuition that someone is transferred with the cerebrum can perhaps be countered by a similar switch from the cerebrum-donor's point of view to the cerebrum-recipient's point of view.[42] Suppose you are told that you will receive a new cerebrum, perhaps in order to save you from an inoperable tumour in your current cerebrum. The new cerebrum cannot be wiped of old memories,

[41] Williams (1970). [42] Snowdon (2014: 11.4).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

intentions etc. without damaging it beyond repair. So not only will you lose your memories but you will gain new apparent memories. Happily however, your loved ones and doctors will rally round, help you re-learn your real name and history, get you used to your house again, reintroduce you to all your favourite music, food, and hobbies.

To the extent that it is easy to go along with this description of the case, we have generated a contrary intuition to the usual intuition that anyone moves with a transplanted cerebrum. In principle it would seem similar to receiving a kidney or liver transplant.

How powerful is this response? Here are two critical comments. First, the counter-intuition is certainly not stronger or more stable than the original. If we cease to ignore the history of the cerebrum then it lapses. Imagine after the operation come the protests: 'no that's not my name! … you're not my family! … this is a terrible mistake … I'm not who you people think I am'. The re-education program begins to look at best like wishful thinking on the part of the relatives, and at worst like the most sinister gas-lighting. (Imagine the doctor's instruction: 'in the interest of the program, the patient must be kept away from telephone and email, or else he may try to contact strangers asking to be rescued.')

Another point is that a cerebrum transfer argument against animalism need not involve a case with a recipient animal at all. Suppose that a cerebrum is removed intact from an animal which is then destroyed. The cerebrum is placed into a vat of nutrient fluids where it is stimulated into activity intrinsically matching its usual intracranial activity when one is conscious. It seems possible that one could continue to exist in such a situation, perhaps even continuing a particular chain of mental activity from start to finish. But, arguably, the animal would be gone. Since you and the animal differ with respect to your existence in this possible situation, you are distinct things. But the intuition here can't be countered by presenting the case from the point of view of the recipient before the operation. There was no recipient, or anyway no recipient with a point of view (an unoccupied vat of nutrient fluids has no point of view).

Even if it can be deprived of its intuitive pull, there may be a more theoretical problem with denying the cerebrum transfer verdict. Johnston asks us to focus on the cerebrum alone, whether settled in a vat or in transit to a recipient animal. It could sustain a conscious subject of experience. But if this conscious subject is *not* you, then where did it come from? What is this 'remnant person'?[43]

No answer to the 'remnant person problem' seems very satisfactory. Has destroying the non-psychological parts of an animal brought it into existence

[43] Johnston (2007). This problem moved Parfit (2012) to reject animalism.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

a new subject of experience? It sounds odd to say that you could bring a new subject into existence just by removing parts from another subject.[44] Perhaps then it existed all along, as part of the animal, but only came to be a subject when excised from the animal. But this also sounds strange: why should fleshy surroundings prevent it from experiencing, while vat fluid surroundings allow it to experience? Perhaps then it was experiencing all along, even when contained within the rest of the animal. But if the animalist says this then they face just the same kind of 'too many thinkers' difficulties as the neo-Lockean alternative. With what right could you claim to be the animal rather than the undetached potential remnant subject? When you think you are an animal, it thinks wrongly that it is an animal. How do you know you're the lucky one?

We will revisit the problem of undetached proper part thinkers in Section 4. But for now, enough has been said to make clear that there are challenges for the denial of the transplant intuition as well as for the denial of animalism. This motivates the investigation of the remaining animalist option: the option of arguing that the animal would go with the cerebrum.

3.4 Animalism with Psychology

Why not use the intuitive verdict that we would go with the cerebrum as evidence about what is possible for animals of our kind? Since they are we, these animals go with the cerebrum.

This might be a convenient thing to say but does it cohere with any plausible view of the persistence conditions of animals? It would flow from the view that animals have the persistence conditions attributed to us by neo-Lockeans: that animals continue to exist just in case they continue psychologically.[45] But first, psychological continuity seems to have nothing to do with the persistence of many biological organisms. Plants and marine sponges for example continue to exist without any psychology. So the view would seem to require some special pleading on behalf of animals. Second, even restricted to animals, the view would face problems with foetuses and persistent vegetative cases. Setting aside whether *we* would persist in such cases, it seems plausible to say that the animal at one's location was once a non-psychological foetus and could one day become vegetative if severely brain damaged. Psychological continuity does not seem to be necessary for the persistence of these animals.

Many animalists have instead adopted a purely 'bodily' (i.e. non-psychological) view of the persistence of the relevant animals: these animals continue to exist just

[44] But we will revisit 'creation by detachment' in the final part of Section 3.

[45] McDowell (1997) suggests something like this position. Sharpe (2015) defends the combination of animalism with Lockean persistence conditions. Bailey (2015) reviews some ways of combining animalism with 'non-biological criteria of identity over time'.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

in case their non-psychological biological functions continue: breathing, circulation of blood, digestion, and other processes of metabolic self-maintenance.[46] This bodily view coheres with intuitive verdicts about foetal and vegetative cases of animal persistence without psychology.

But it is important to emphasize that a bodily theory of animal persistence is not entailed by these verdicts. These verdicts only require that psychology is not *necessary*, and that in its absence non-psychological continuities can be sufficient for an animal to persist. So it is not too difficult to devise a hybrid view which simultaneously respects both these verdicts and the cerebrum transplant verdict: non-psychological bodily continuity while not necessary is sufficient for the animal to persist, and cerebral psychological continuity while not necessary is sufficient for the animal to persist.

An immediate difficulty for this view is that in the cerebrum transfer case as originally presented (unlike the isolated brain in a vat case) there is in effect a *fission* of these kinds of continuity, and not one without other. There is something which exhibits mere vegetative continuity and there is something which exhibits psychological continuity. If these continuities are each sufficient for animal persistence, then the animal goes *both* ways. But we don't think of an animal as something which could come to be wholly located in separate places like a universal, and nor do we think of an animal as something which could come to be partly located in separate places like a scattered tea-set. If these are not possibilities for an animal, then the view is incorrect.

In response one could amend the hybrid view in either of two ways. According to a 'no-branching' variant, psychological and bodily continuities are each sufficient for animal persistence in the absence of the other, but when both obtain, the animal ceases to exist. But obviously this version would not serve the purposes of the animalist who wishes to preserve the intuition that we would go with the cerebrum when it divided away from other parts of the animal. A more attractive variant is a 'best candidate' hybrid view: psychological and bodily continuities are each sufficient for animal persistence in the absence of the other, but when both obtain, the animal continues with its psychology. Psychology dominates. This version would allow the animalist to preserve the intuition that we would go with the cerebrum.[47]

Despite its attractions the hybrid view developed in this way faces challenges.[48]

First, it appears to be unprincipled. The theory has a disjunctive or gerrymandered appearance, as if cooked up simply to fit otherwise opposing intuitions

[46] van Inwagen (1990) and Olson (1997). [47] Madden (2016) and Noonan (2019).
[48] See (Olson 1997: 114–119; Olson 2015).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

about ourselves. It does not appear to flow from any biologically realistic view of animals. From the point of view of biology, isn't a cerebrum just another small part of an animal, like a liver or kidney? No one thinks an animal would go with its liver if it was divided away from the rest of the animal. Why is the cerebrum any different?

Second, the view shares the prima facie strange consequences of other views with a 'best candidate' structure. A similar hybrid view arises from attempting to reconcile our intuitions about the famous Ship of Theseus. It is quite plausible that the continuously functioning ship which has had its planks gradually replaced is the original ship, whether or not a ship is later built from the old planks, On the other hand, if the planks had just been scattered and then put together again without being replaced to sustain a continuously functioning ship, then it is quite plausible that the original ship would have been rebuilt from the planks. A theory which respects these intuitions says that functional continuity and continuity of material parts are each sufficient for ship persistence in the absence of the other, but when both obtain, the ship goes with the path of functional continuity.

It's been objected that this theory makes the existence of ships unacceptably dependent on 'extrinsic' factors.[49] Consider the path of the planks in the case where there is no functioning ship. A ship persists along this path according to the theory. But this path seems to be intrinsically replicated in the case where there is also the functioning ship, which could be hundreds of miles away by the time the planks are fully reassembled. Yet according to the theory, no ship persists along the planks' path in this case: we have to view the ship made from the scattered planks as something newly created. But how could what happens somewhere else make the difference between whether the process of scattering and reassembling sustains one ship or brings into existence a new ship? If a process with a certain here-and-now intrinsic character sustains a single thing of a certain kind, then shouldn't any process of the same character also sustain a single thing of that kind? How could what happens somewhere else prevent it from doing so?

The same general worry applies to the best-candidate hybrid theory of animal persistence. In the case where an animal has its cerebrum simply destroyed, the theory says that the animal persists in a vegetative state. But this path seems to be perfectly replicated in the case where the cerebrum is preserved and implanted. According to the theory, the animal does not persist along the vegetative path in this case. We have to view the vegetative animal as something newly created. But how could what happens somewhere else make the

[49] Wiggins (2001) and Noonan (2019).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

difference between whether the vital processes ongoing here sustain one animal or bring into existence a new animal? If processes of a certain intrinsic character sustain a single animal in the one case, shouldn't any process of the same character sustain a single such animal? How could what happens somewhere else prevent it from doing so?

Further strange consequences centre on the recipient of the cerebrum. If the donor animal and not just an organ moves with the cerebrum then what happens to the recipient animal? Are there now two animals in the same place? Or has the recipient animal somehow been destroyed, by implantation of a perfectly healthy organ? Neither option sounds palatable.

The best-candidate hybrid theory of animal persistence needs to say something to address these challenges. A promising way to address the first is to look more closely at the nature and persistence of biological organisms generally.

3.5 Organisms and Persistence

We start with some general comments about persistence and then go into some detail about biological organisms.

The continued existence of an ordinary concrete object along a certain trajectory or path is explained by causal processes intrinsic to that path. The character of these causal processes will vary between kinds of object. Some of these 'processes' will be quite boring or static connections; others will involve more dynamism and change. For example, the continued existence of a rock is explained by a stable crystalline structure simply perpetuating itself from one moment to the next. The continued existence of a star requires continuous nuclear explosions.

For a kind of concrete continuant with complexity, more than one form of internal causal connection is likely to be relevant to the persistence of its specimens. For example when a computer persists there are internal causal connections ranging from the boring perpetuation of the gross structure of over time, to more dynamic causal dependencies between inputs, memory-states, processing, and outputs over time. In cases such as these, when a kind is associated with a diverse set of internal causal connections, the persistence of an individual specimen of the kind will require that some sufficient subset (not necessarily all) of these connections hold. That is to say, a 'cluster theory' of persistence is to be expected for complex kinds of continuant.

What is a biological organism? An abstract description goes as follows. Biological organisms are self-maintaining individuals, with mechanisms of self-maintenance realized by very complex chemical processes involving carbon and energy exchange. They have the capacity to be at least a partner in the

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

reproduction of similar things, and the historical explanation of the existence of species and higher taxa of organisms refers to the fact that the reproduction of organisms involves heritable variation of differentially adaptive traits.[50]

What explains the persistence of a biological organism? A natural starting point is to focus on its biochemically realized mechanisms of self-maintenance. These make biological organisms excellent examples of concrete continuants as characterized. Their persistence involves rich internal causal connectedness. The operation of a range of mechanisms within the structure of an organism causally explains its structure and existence at subsequent times. Taken collectively, the operation of these self-maintenance mechanisms could be called the *life* of the organism. The life of an organism is what generates the internal connections constitutive its persistence.

To say this is not to rule out an organism's continuing to exist when it is no longer alive. Immediately after the operations of self-maintenance stop (as long as the death has not been too violent) we will still have, though perhaps only for a short time, a biochemical structure which is causally explained in its very fine details by the earlier operation of those mechanisms. The cross-temporal internal causal links generated by life may in principle extend a little way beyond the cessation of life.[51]

'Biochemically realized self-maintenance mechanisms' remains a highly schematic element in the characterization of the persistence of biological organisms. The availability of a single schematic characterization of the persistence of biological organisms does not show that there are not significant differences between the persistence conditions of different kinds of organism, any more than the generic applicability of the schematic notion of 'internal causal connectedness' shows that all concrete continuants have the same persistence conditions. If one were to draw the latter conclusion then one would be liable to unsound reasoning on its basis. For example: all concrete continuants have the same persistence conditions. Rotation of water is irrelevant to the persistence of tables. Therefore, rotation of water is irrelevant to the persistence of waterspouts. There might be variation in substantive persistence conditions across things to which a single schematic persistence condition applies. This is not to say that the substantive persistence conditions of biological organisms do in fact vary. The question of whether there are, or are not, significant differences between the biochemical self-maintenance mechanisms of different kinds of biological organism is not a question which can be settled a priori. We need to turn to some empirical details.

[50] Compare NASA's working definition of life: 'a self-sustaining chemical system capable of Darwinian evolution'. See (Joyce et al. 1994).

[51] Cf. Ayers (1991: 184).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

3.6 Varieties of Biological Self-Maintenance

A single-celled organism such as a bacterium has a well-defined membrane boundary which conceals what one biophysicist has called a 'molecular storm' (Hoffman 2012). Chemical reactions within the cytoplasm sustain the structure and boundary of the cell in the face of its thermodynamic disequilibrium with its surroundings. We find a finely-tuned balance of 'constructive' anabolic processes, which draw upon energy to create complex organic molecules like proteins and lipids, and 'destructive' catabolic processes which break down molecules to release energy and simpler chemical products.

The picture of a crisply individuated entity which perpetuates itself by basic metabolic reactions captures an important aspect of unicellular life. But as a model for the nature and persistence of biological organisms generally, it is liable to mislead in at least two ways. First, it overlooks significant self-maintenance mechanisms characteristic of even unicellular life. Second, it fails in some ways to scale up to multicellular life. We will take these points in turn.[52]

The first thing to observe is that even a bacterium *regulates* its metabolism. It has, in addition to the basic mechanisms which construct and break down chemicals, higher-order mechanisms which, depending on circumstances, control or speed up these metabolic reactions, usually by means of enzymes. Sometimes regulation involves switching on new metabolic pathways. For example, the *lac* operon system in *E. Coli* bacteria responds to low levels of environmental glucose and high levels of lactose by activating a lactose-metabolizing pathway that would otherwise be too costly to maintain (glucose is preferred and normally available). It is to some extent a terminological matter whether to count this kind of higher-order regulation of metabolism as itself part of 'metabolism', but either way it is a biochemically realized mechanism whereby the bacterium maintains itself across time.

The example is helpful for two reasons. First, it illustrates that self-maintenance mechanisms characteristic of a kind of organism can operate over different *time scales*. The state of a living bacterium at an instant *t* will have a cluster of partial internal causal explanations which vary in their temporal proximity to *t*. One causal explanation is that certain fast metabolic reactions had reached intermediate stages just one *microsecond* before *t*. Another causal explanation is that the lactose metabolism system switched on *a few seconds* before *t*. If one is to picture the cross-temporal causal links involved in the self-maintenance of the bacterium then one

[52] The rest of this section is influenced by Godfrey-Smith's campaign to get philosophers to reconsider whether there is such a sharp and significant distinction between metabolism and cognition (2016a, 2016b). Other writers in theoretical life sciences have put related ideas in very bold terms: 'Living systems are cognitive systems, and living as a process is a process of cognition'. (Maturana & Varela 1980: 13)

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

should not only picture chains of very short threads. Longer threads are overlaid too, spanning longer periods of time. As the pattern of threads in a twill fabric can repeat at different lengths, so the internal causal history of an organism is bound together at different time scales.

A second important illustrative feature of the *lac* operon system is this: it is an example of self-regulation that involves adaptation to changing *environmental* circumstances; here ambient sugar levels, rather than directly to internal metabolite levels. Bacteria regulate themselves, not only by monitoring and adjusting to interior conditions, but also by monitoring and adjusting to exterior conditions. Another nice example of this sort of adjustment, also from the study of *E. Coli* bacteria, is *chemotaxis*.[53] The flagellae of the bacteria will be more likely to continue to move coherently if environmental chemical conditions are changing favourably (glucose is increasing, or toxins decreasing, for example). But if conditions are not changing favourably, then the flagellae will be more likely to move at random, causing the bacterium to tumble. The causal effect of this regulatory mechanism, like the *lac* operon system, is to maintain the bacterium in existence over time as its environment changes. Here the mechanism can produce threads of causal dependence over the scale of minutes, rather than seconds or milliseconds.[54] A further difference is that the mechanism achieves this regulation by causing the motion of the organism as a whole rather than just the production of chemicals.

A simple image of unicellular life as a neatly boundaried metabolic storm, then, overlooks self-maintenance mechanisms which operate on longer time scales than basic metabolic reactions. Some of these involve sensitivity and response (chemical or motile) to changing exterior conditions. When we consider larger, eukaryotic, unicellular organisms, these longer time-scale regulation mechanisms diversify further. A *protozoan* for example ingests and digests other organisms, whereas some unicellular green *algae* display flagellar movement towards light to facilitate photosynthesis. Neither do both of these things. So, between and within even unicellular organisms, we find a diversity of self-maintenance mechanisms.

What about the self-maintenance mechanisms of multicellular organisms? In some respects the unicellular organism is a good model, in other respects it is not. A unicellular organism has a well-defined semi-permeable membrane boundary to prevent movement towards thermal equilibrium with its environment. But a multicellular organism does not have, and does not need to have,

[53] For a philosophically oriented comparison of the *lac* operon system and chemotaxis, see (van Duijn et al. 2006).

[54] See (van Duijn et al. 2006: 164) on the time scale of the slower 'memory'-like methylation pathway involved in *E. Coli* chemotaxis.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

such a clear boundary. It maintains conditions for cellular metabolism to continue within its various differentiated cells, and it maintains an overall integrated structure; but multicellularity brings with it a new kind of unclarity about which things are included within it and which things are not. The pervasive phenomenon of collaboration between eukaryotic cells and bacterial cells within multicellular organisms makes the point.[55] The activity of animal gut bacterial flora, needed to break down certain carbohydrates, could be considered within the 'metabolic boundary' of the multicellular organism. But it is not clear, and there would seem to be little point in legislating one way or another. The symbiotic *vibrio* bacteria essential to the bioluminescence mechanism in some squid species are another illustration. Although the multicellular collaboration is composed of genetically different cells, the structure which they compose exhibits division of cellular labour, sharing of energy and nutrition, intercellular signalling, and an overall functional structure which maintains itself and enables the collective to move as a unit. All this without a sharp thermodynamic boundary like a cell.

But even if the neatly encapsulated metabolic storm limps as a model for multicellular life, there are of course continuities with unicellular life. Mechanisms of self-regulation by monitoring and response to *exterior* conditions take many forms in multicellular life. Sensing and adjusting to ambient temperature is one example. However, the example of bacterial chemotaxis is a primitive analogue of multicellular self-regulation mechanisms which are of special interest in the present context. In the 'Cambrian explosion' when, very suddenly by evolutionary standards, all the main animal phyla emerged, the capacities for sensing and motion towards nutrients and away from threats became very sophisticated very quickly. The development of *nervous systems* allowed for quick coordinated contractions of large bodies, with articulated limbs, pincers, and jaws. Complex eyes provided detailed information about the environment of the organism, in the light of which courses of predatory or evasive behaviour could be quickly compared and selected. It is a reasonable hypothesis that adapting to an environment containing predator and prey organisms with these sophisticated mechanisms created a positive feedback effect on the rate of their evolution; hence the 'explosion' of animal life forms in this period. But in any case, ever since this period the biochemically realized mechanisms by which an individual multicellular organism sustains itself have included *cognition* broadly speaking: perception, learning, and flexible behaviour. Of course this is not true of every multicellular organism; plants and

[55] See (Dupre & O'Malley 2009; Godfrey-Smith 2012; Pradeu 2012).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

fungi do not have cognitive capacities in this sense. But animal kinds are characterized by nervous-system-based cognition.

These functional developments correspond to anatomical developments. Complex cognition has developed in at least two of the animal *phyla* which emerged in the Cambrian: in *mollusca*, which includes octopuses, and in *chordata* which includes vertebrates. The chordate body-plan is organized around a dorsal nerve cord, which, in vertebrates becomes the central nervous system with a large brain at one end. The brain is a concentrated structure realizing, as well as some non-cognitive mechanisms of monitoring and homeostasis, the very complex cognition which is characteristic of vertebrates.

In *homo sapiens*, capacities to monitor and adjust to a complex exterior environment have reached new heights. The environment of these organisms contains not only ambient temperature and glucose gradients, not only predators and prey with quick sensory-motor systems, but also complex tools and social groups. The causal mechanisms by which these vertebrate organisms adjust to their complex environment distinctively include linguistic communication, causal reasoning, episodic memory, mind-reading, imitation, and other mechanisms of social learning.[56] This increased functional complexity corresponds to an intensification of the vertebrate anatomical trend towards large brains; the human cerebral cortex, the structure most closely associated with cognition, contains some 16 billion neurons, around twice that of our closest primate cousins.

Let us bring together the results of the preceding. Complex continuants persist in virtue of a cluster of internal causal connections characteristic of their kind. Individual biological organisms are highly internally causally connected by the operation of biochemically realized self-maintenance mechanisms. The natural view is that biological organisms persist in virtue of the cluster of internal causal connections produced by these mechanisms. What should now be clear from our review is that while basic cellular maintenance by metabolic reactions may be common to all biological organisms, mechanisms of self-maintenance ramify and diversify from the unicellular level upwards.

Correspondingly it seems right to say that different kinds of organisms persist in virtue of different clusters of causal connections. The cluster of causal connections associated with the kind *E. Coli* may include, for example, the causal connection between the detection of increased lactose at one moment and the activation of lactose metabolism at the next. The cluster of causal connections associated with a multicellular plant kind may include, for example, the causal connection between the opening of leaves at one time, and increased carbon levels at a later time. The cluster of causal connections associated with

[56] See (Heyes 2018) for a recent review of distinctively human cognitive mechanisms.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

a multicellular animal kind may include, for example, the causal connection between visual prey detection at one time and contraction of pincer flexor muscles soon thereafter. For a mature specimen of *homo sapiens* these latter, cognitive, causal connections are greatly increased in variety and number, relative to even other primate species. The cluster of internal causal connections constitutive of its persistence will still include a range of brute 'bodily' connections, of basic physical cohesion, ongoing metabolism, as well as longer-range interior regulation of nutrition, toxins and temperature. But it is a reasonable view that these will not predominate over the vastly complex constellation of psychological connections laid down by cognitive mechanisms. Consider the range of detailed connections involved in the organism's adaptation to its highly specific socio-cultural environment, its numerous episodic memories, diverse fine motor skills, action-planning, syntactic and semantic linguistic competences, perceptual-recognitional capacities, and the range of specific exemplifications of causal understanding required for sophisticated tool-use and other adaptive modifications of its physical environment.

So we reach the point at which we can draw the conclusion that some biological organisms have 'psychological criteria of identity', in at least the following sense: the clusters of causal connections in virtue of which biological organisms of some kinds persist include many causal connections laid down by cognitive mechanisms. For a mature, functional, specimen of *homo sapiens,* the number of such causal connections will be vast.

3.7 Revisiting the Challenges

Since it is plausible that we are animals, the intuition that we would go with the cerebrum gives prima facie grounds to judge that animals of our kind would go with the cerebrum. What we have seen now is that the perspective of theoretical biology does not push back against these prima facie grounds. The nature of animals supports the inclusion of psychology in their persistence conditions. Different kinds of organisms are characterized by different (though overlapping) biochemically realized mechanisms of self-maintenance. In the case of animals and especially *homo sapiens*, these mechanisms include diverse cognitive mechanisms each of which generates a very large number of causal connections between the states of the animal at different times. As a matter of anatomical fact, the cerebrum is the seat of this vast array of causal connections. No other similarly sized part of a mature, healthy, *homo sapiens* would transmit so many causal connections characteristic of this kind of biological organism. To single out the cerebrum as especially significant for the persistence of an individual *homo sapiens*, then, would seem to be the opposite of unprincipled.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

What about the claim that even when the cerebrum-complement is still on the scene, the animal goes with the cerebrum? Nothing in our review of biological self-maintenance opposes this verdict; indeed it may support it. One view is that the mechanisms of self-maintenance which are especially distinctive of a kind of organism get special weight in determining the persistence of that kind. In the case of animals, we have seen that these are cognitive mechanisms. However, the theorist need not take this inegalitarian view. When the cerebrum and cerebrum-complement are separated, there is a degree of branching or fission of causal connections generated by characteristic self-maintenance activity. A case could made that in a mature, healthy, *homo sapiens* the complex cerebrum simply sustains *more* of these detailed connections than the cerebrum-complement. Consider not just the impressive range of generic cognitive mechanisms, but the fact that in a mature human each of these mechanisms has generated a vast number of specific causal connections sustained by the cerebrum.[57]

Of course the cerebrum-complement is more massive. But we should not count the human characteristic causal connections sustained in the two parts respectively by counting atoms, or by weighing flesh. The connections hold between instantiations of higher-level properties ascribable to animals, not between states of atoms. A naturally fine grain of individuation of such higher-level properties may well favour the cerebrum. To take one example, it is explanatorily natural to individuate states of knowledge by propositions known. So there are at least as many cognitive connections as there are propositions knowledge of which is retained. In contrast, the non-psychological connections in the cerebrum-complement are arguably relatively sparse. There are immensely complex organic chemical processes ongoing at a lower level of realization, but the higher-level properties of an organism which are thereby realized amount to a rather simple life: breathing, snoring, and excreting.

Even if the view is not unprincipled, it still needs to face up to the extrinsicness concern about theories with a 'best candidate' structure. Recall the issue: in the case where the cerebrum is simply destroyed, the animal persists in a vegetative state in virtue of sufficient (non-psychological) connections. In the cerebrum transplant case, similar connections fail to sustain a single animal. The animal goes with the dominant (psychological) connections. So the cerebrum-complement must be a new (vegetative) organism.

The concern about 'extrinsicness' here is elusive. It needs to be distinguished from other spurious objections. First, it might seem that the necessity of identity and distinctness is violated: a particular which is distinct from the animal

[57] This need not be true of an early fetus, or a human with a severely damaged cerebrum. In such cases the cerebrum-complement is more likely to be dominant.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

(the vegetative remnant) would have been identical to the animal if the cerebrum had been destroyed rather than transplanted. But this objection is misguided. If the cerebrum had been destroyed then this vegetative remnant would simply not have existed. Instead the original animal would have been in its place, reduced to an intrinsically similar vegetative state. If things are distinct or identical then they are necessarily so. Such facts are not contingent on anything. *A fortiori* they are not contingent on whether certain processes occur or not.

Does the theory have the consequence that the animal's persistence is not determined by the way it is intrinsically? No. On the current theory, the asymmetric division is a wholly intrinsic change in the original animal at a certain time. Being divided along a certain line is a different intrinsic change to shrinking by cerebrum loss. Thus the path along which the animal will persist at one moment is fixed by its intrinsic character at an earlier moment. It is not determined by anything outside of the animal, nor by subsequent events. There is no failure of the intrinsic or 'immanent' determination of a concrete thing's persistence from one moment to the next.

There are some wilder versions of 'best candidate theories' which hold that the original thing only goes along the stronger branch if the stronger branch does not turn out to terminate much earlier than the weaker branch: if it does, then the original thing either 'jumps' to the weaker branch, or somehow comes to have been on the weaker branch from the moment of fission in virtue of what would happen later.[58] On either theory the here-and-now condition of the original thing at the moment of fission would fail to fix its path of persistence to the next moment. The view under consideration does not have this strange consequence.[59]

So on the present view everything in the story *persists* in virtue of the way it is intrinsically. But perhaps it will be objected that something is *created* in an objectionably extrinsic way. The matter-filled region of space-time which first contains the animal and then the distinct vegetative remnant left behind by the cerebrum is intrinsically similar to a matter-filled region of space-time which contains a single persisting organism whose cerebrum has simply been destroyed. How can the first region but not the second involve the coming into existence of a new thing?[60]

The conviction that this is not possible may rest on a myopic focus on the paradigm of creating of material things by bringing matter into a region.

[58] Versions 2–4 of Nozick's 'closest continuer theory' (Nozick 1981: 42–43) do not respect the intrinsic determination of persistence.

[59] See (Hawley 2005) for criticism of such views.

[60] Noonan's favoured formulation of an 'only-x-and-y principle' (2019: 40) is designed to rule this out. This drives him to the surprising position that the vegetative animal must have existed all along: so two animals were in the exactly same place before the operation.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

A bounded material thing can also be created by detachment or dispersal of matter from around a matter-filled region. For example, consider the fact that a figurine can be created by carving away from a large block of wood. Suppose that such a figurine is created on Monday and persists to Wednesday. In a second situation a figurine is created from the same block of wood on Tuesday instead. Still, there is in the second situation a figurine-shaped region of cellulose-filled space-time that extends back to Monday. It internally matches the cellulose-filled region of space-time swept out by the Monday–Wednesday figurine. And yet in this case a figurine only comes into existence half-way through the period.[61] How could this be?

There is no mystery once we remember that ordinary kinds of material objects are not themselves matter-filled regions of space-time but essentially have a *contrast* with their surroundings. That's why they can be created by detachment. Detachment creates a new contrast. The new vegetative organism is created by detaching it away from the other (smaller) part of the animal. If there had been no other parts of the animal (i.e. if the cerebrum had simply been destroyed) then there would have been nothing from which it could detach and no new boundary would be created. The animal would just lose its cerebrum. Similarly, in the Ship of Theseus case a new ship is created by detaching planks from the persisting ship. If there had been no continuously functioning ship there would have been nothing from which it could be detached, and thus no new boundary would be created. The ship would have moved with the planks.

Of course detachment is not the usual way of creating animals. But it is useful to note that the biological world seems to contain organisms which are actually created in this way. Horticulturists propagate plants by taking *cuttings* as well as growing from seed. A piece is detached from a parent plant and if suitably watered and nourished will itself come to grow roots and flourish. This is a case of asymmetric fission. The cutting is usually regarded as a new specimen, while the parent plant persists, slightly reduced. In the cerebrum transplant case, on the view under consideration, it is the larger 'cerebrum-complement' which is the new cutting, and the smaller cerebrum-shaped thing which is the parent.

Similarly, it may be easy to overlook the fact that the destruction of a material thing need not involve dispersal of matter in the object's region but can be brought about by addition of matter around that region.[62] A snowball caught in an avalanche is one example. Horticulture, again, provides others. The fate of

[61] *Pace* Noonan's only-x-and-y principle.

[62] Material bodies are exactly converse to *holes* in these respects. Holes are paradigmatically created by deleting matter from a region, but can also be created by surrounding an empty region with matter. Holes are paradigmatically destroyed by filling an empty region with matter, but can also be destroyed by deleting matter surrounding an empty region.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

the recipient animal is nicely modelled by the *grafting* of a small fragmentary plant onto a larger plant. This is plausibly a case of asymmetric fusion in which eventually only the dominant plant survives, albeit with some characteristics inherited from the graft. Again, in the cerebrum-transplant case, weight of flesh is a poor guide to dominance in kind-characteristic causal connectedness: it is the larger vegetative cerebrum-complement which is grafted onto the small, briefly cerebrum-shaped animal, and which is eventually absorbed by this dominant organism.

To sum up: attention to the biological world, botanical as well as zoological, weakens the usual kind of persistence-based objection to the view that we, the objects of our self-representations, are identical to certain animals. It does not seem to be an unprincipled or strange view that the persistence of certain animals can be secured by psychological continuity.

4 Multiplicity

We've now seen that there is mileage in the view that some animals have partly psychological persistence conditions, and that this undercuts the strongest persistence-based argument for distinguishing ourselves from animals in the first place.

This section considers some other important arguments for distinguishing ourselves from animals. There are actual or nearly actual cases in which there appear to be more thinkers than there are animals. Such cases include two-headed conjoined twins, multiple personality disorder, and so called 'split brain' cases. These cases generate important arguments for distinguishing ourselves from the animals in our vicinity.

But we will begin with a reason for thinking that even in ordinary, non-pathological, cases there are more, perhaps many more, 'owners' of mental phenomena than there are animals. Collectively these arguments offer some support to the minimalist view that we are not whole animals but brains or brain-parts of animals. However we will also present a response on behalf of the view that we are animals. This in turn suggests various ways of looking at the twinning and split-brain cases.

4.1 The Problem of Thinking Parts

Consider the undetached head of a normal awake human being, still attached to the rest of its parts. Now consider a possible case in which a human is drastically mutilated down to just its head. Suppose that this mutilated human exactly matches, if only for a moment, the internal neural configuration of the attached head of the normal awake human being. It is plausible that the mutilated human

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

thinks and experiences. But now, from the 'internalist' assumption that, if a pair match exactly in their internal material configuration, then the pair match in thinking and experiencing, it follows that the undetached head of the human being also thinks and experiences: it is 'in itself' just like a possible small psychologically endowed human; therefore it too is psychologically endowed.

By reasoning in the same way about various possible amputee thinkers, it can be argued that every ordinary human being contains a host of undetached brain-overlapping parts, all thinking and experiencing: the undetached brain, the undetached head, the undetached upper-half, the undetached left-leg-complement, and so on indefinitely. All are internally like possible thinkers, so each is a thinker.

On the face of it this is an absurd result.[63]

However a no-bearer theorist might be relatively relaxed. If there is never a particular object upon which an experience is ontologically dependent, then all of these brain-overlapping objects can be regarded as containing literally the same experiences within themselves. Seen in this way, the case is not clearly more paradoxical than the fact that a certain group of shiny jewels is simultaneously contained in all of the following nested locations: a chest, a room, a tower, London, England, and Europe. To say so is not to commit to a multiplicity of jewels.

What would this multitude of thinkers mean for the question of what we are?

It might have a no-self result. Perhaps nothing makes one of these thinkers in particular the reference of the self-representations which they all contain within themselves. In which case there is nothing in particular that we are.

As noted in Section 2, it does not seem plausible that there is a sortal-concept associated with self-representation which would promise to focus self-representation on any thinker in particular. However the epistemic charity approach to reference suggests some other options: it remains plausible that the multiply contained self-representations are specially sensitive to reflexive channels of knowledge of the organism in particular. So, despite the multiplicity of nested mental systems, the organism may well be the knowledge-maximizing assignment of reference to self-representations. The truth of animalism may be robust in the face of a multiple thinkers result.

An intriguing alternative possibility is that a charitable approach to reference induces a kind of flexibility about the reference of self-representations. For example, if the self-representations include the belief that one is a head, then the assignment of the head as reference will do better on the score of truth than the

[63] Johnston's 'personite problem' (Johnston 2016) is a parallel problem of multiple thinking *temporal parts* rather than spatial parts. It deserves less attention because the existence of temporal parts of human beings is far less intuitive than the existence of spatial parts of human beings.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

assignment of any non-head such as the whole organism. If this superiority of fit makes it the case that the head wins out as the reference of self-representation then the self-representation 'I am a head' will in fact turn out to be true. In a sense it could be up to us what we are.[64]

Although this view may have a certain political appeal, it will need to deal with the fact (recall twin-mother in Section 2) that the best-fitting assignment is not necessarily the object with which one has the right kind of epistemic contact for reference. It is not clear how strange opinions about one's spatial extent would count as knowledge even if they are true.

Moreover, even if the question of what we are can be argued to have a positive answer one way or other, those who conceive of an experience as by its nature the experience of a particular bearer should be more alarmed by the result that there are multiple experiencers. If the experience of an experiencer has some special relation to that experiencer alone, then the experiences of different experiencers must be distinct. On this view, the multitude of experiencers corresponds to a multitude of experiences, a multitude of entire mental lives running in parallel.

It is time to give the assumptions of the initial argument for multiple thinkers more scrutiny.

Recall in Section 3 that a 'too many thinkers' problem was dissolved by undercutting the reasons for believing in the rival non-animal candidate thinker: if no distinct Lockean person needs to be posited in order to respect the brain transplant intuition, then the too many thinkers problem goes away. Can the thinking parts problem be dissolved in the same fashion, by denying the existence of the rival candidate thinkers?

It seems unlikely. In this case, the non-animal candidate thinkers have not been introduced on the basis of any special thought experiment. It is fair to say that the existence of some of the brain-overlapping parts is questionable: it is not obvious that the act of abstracting away from an organism's left foot succeeds in selecting a real concrete object which has all of the organism's parts but for the left foot. One might deny that such an undetached part of an organism exists, or more subtly, accept that it exists but hold that it is a kind of abstract spatial object, like the equator or the northern hemisphere. On the latter view, the internalist principle would be denied application: no abstract object is 'in itself' like a concrete human thinker.

But it would be more difficult to deny that an undetached head is a real concrete object. Such a thing is salient to us without any unusual descriptive

[64] Kovacs (2016) explores this possibility. For a kindred self-determination view see (Kurstal 2022).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

stage-setting, and it has some causal unity of its own, its parts moving together rigidly and to a degree independently of the rest of the organism.

Some theorists have denied the existence of any composite objects besides organisms. This eliminativist position would make the problem of thinking parts go away: the organism wins the title of owner of the local thoughts and experiences by default. It is the only candidate in the vicinity.[65] The eliminativist position is not to be dismissed lightly in the light of its puzzle-dissolving power. But at least at first glance it is profoundly unattractive.

Are there ways of denying that there are multiple brain-overlapping thinkers that do not deny the existence of undetached heads and brains?

4.2 Bearers: The Minimal Realizer Approach

In Section 2, the debate about whether experiences have bearers was largely approached 'from the inside': does experience seem to have a bearer from the point of view of introspective reflection? Now it is time to look at two principled ways of introducing bearers 'from the outside'. Each promises to select a single bearer from among the animal and its various undetached parts. We will call these the minimal-realizer approach and the teleo-functionalist approach. Roughly, the former 'zooms in' to the brain. The latter 'zooms out' to view a larger functional system in relation to its environment.

To explain the minimal realizer approach we can pursue the analogy of a jewel located in multiple nested locations. Special among these locations is the smallest jewel-shaped region which not only wholly contains the jewel, but is also wholly filled by the jewel. This region, which has exactly the same shape and size as the jewel, we can say the jewel *occupies*. The jewel is also located in larger spatial regions, but derivatively so, in virtue of the fact that these larger regions contain the region which the jewel occupies.

In parallel, we can suppose that for each experience, among the range of nested wholes the intrinsic state of each of which suffices for an experience of that kind, there is a smallest part the intrinsic state of which suffices for an experience of that kind. This is the minimal-realizer of the experience. There is a smallest part of an animal 'having' an experience, such that were the animal cut down to that part, and the part stimulated to the intrinsic state it is in when still part of the animal, then an experience of the same type would occur.

It might be said that larger wholes which have this smaller part count as experiencing, but only in a derivative sense, in virtue of containing the thing which is in a more basic sense experiencing. A suggestive comparison is a car with a dent in it. The whole car can appropriately be said to have a dent, but this

[65] Van Inwagen (1990).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

is in virtue of a smaller part of its surface having a dent in a non-derivative sense. This smaller part of the car is, as it were, wholly dented. Or compare woodworm infestation: a house has woodworm, but this is in virtue of certain timbers of the house having woodworm. These are examples of events, and more static modifications, which inhere primarily in parts of certain wholes.

However the possibility of identifying a minimal realizer of an experience does not by itself show that experience is appropriately ascribed to that thing. Suppose that a human animal goes for a walk. An otherwise similar human animal missing a finger on one hand could go for a walk of the same kind. It follows that the minimal realizer of walking must be something smaller than the whole animal. But it hardly seems to follow that the animal goes for a walk only derivatively, in virtue of this smaller part of the animal going for a walk in some more basic sense. Walking seems primarily to be an activity of a whole animal. We wouldn't call the activity of the minimal realizer 'walking' at all.

The minimal realizer theorist may retort that this just reflects the way we talk. Since we don't ordinarily notice the minimal realizer of walking, we only ascribe walking to the larger and more salient boundaried entity. But this is not a metaphysically perspicuous practice. We could change the way we talk.

But another critical point about the minimal realizer approach to selecting a bearer of experience should be made. It might be expected to select the brain as the bearer of a given set of mental phenomena. But it is actually likely that distinct experiences in a single mental life will have slightly different neural realizers: an auditory experience need not involve the visual cortex; a visual experience need not involve the auditory cortex. At a time, and over time, the experiences in a single mental life are likely to have distinct minimal realizers within the brain.[66] At least as a solution to a problem of many thinkers, the approach is unlikely to satisfy. It rules out multiple large thinkers at the cost of finding multiple small thinkers.

But if this were the only principled way to identify the bearer of an experience then perhaps we should learn to live with the picture. There is a crowd of small things bearing the experiences associated with the life of one animal. Nothing immediately follows about what *we* are, since it is a further question what is the reference of our self-representations. The self-representations among these experiences might all refer to the animal as the most charitable assignment. Some philosophers might insist that they overridingly intend to use 'I' to refer to whatever is non-derivative bearer of their thoughts and experiences.[67] But in that case, their

[66] Lee (2017) embraces this result in support of the position that elements in a mental life have no single bearer and in that sense are 'selfless'.

[67] Parfit (2012: 21) decides to use a special pronoun 'Inner-I' to refer to the part of the animal which 'directly' thinks his thoughts.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

'I' reference is likely to fail given that the mental phenomena introspectable from within a mental life are likely to have multiple non-derivative bearers.

In the absence of any other principled theory of what it is to be the bearer of mental phenomena, the simple view that a mental life has just one bearer which is also the reference of its self-representations is likely to be viewed as no more than a failure to get messy metaphysical reality into sharp focus.

4.3 Bearers: The Teleo-Functionalist Approach

There is another principled theory. According to broadly functionalist theories, to be a bearer of mental phenomena is to be something with a certain systematic structure. Mental phenomena are individuated in terms of their causal role in mediating between inputs, outputs, and other mental phenomena. As Shoemaker (1984, 1999) has emphasised, standard descriptions of this causal role appear to presuppose the identity of the subject of the state in question. On the input side are dispositions to be caused by sensory inputs to, and by other mental states of, the very thing which has the state being described. On the output side are dispositions to cause other mental states of the very same thing, and to cause behaviour of the very same thing. So a functionalist metaphysics of the mind makes it plausible (contra the no-bearer theory) that an experience or other mental episode is by its very nature related to an owning system.

Here is the associated theory of what it is to be a bearer of mental states:

> (F) *x* is a bearer of mental states if and only if *x* has states which are in fact disposed to causally mediate in the right way between inputs of *x*, outputs of *x*, and other states of *x*

This gives a less promiscuous way of counting bearers of mental states than the minimal realizer approach. Distinct mental states may inhere in single bearer even when their minimal realizers are distinct, so long as those states are causally interrelated in a certain way. A bearer of mental states is a kind of functional unity.

However this approach does not obviously help with the original problem of thinking parts. Consider the undetached brain. It would seem to mirror in microcosm the psychological causal structure of the whole animal. It has 'inputs' which are not changes to eyes and skin, but more proximal stimulations of afferent nerve entries to the brain. Likewise its 'outputs' are not limb movements, but efferent nerve signals exiting to other parts of the animal. Functionalism is standardly intended to allow that thinkers of different kinds could count as bearers of mental states. It is hard to see why the animal, the undetached brain, the undetached head and so on, should not *all* be counted as meeting condition (F). They are nested but numerically distinct functional unities.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

In order to see a way forward, we should note that Shoemaker's functionalism is an example of what Godfrey-Smith (1996: 13) has called 'dry functionalism': it defines mental states in terms of *de facto* causal dispositions, in terms of how in fact the state tends to causally interact. The alternative involves, in Sober's words 'putting the function back into functionalism' (Sober 1985): *teleo-functionalism* defines mental states in terms of what it is their *function* to do. Here a notion of 'proper function' is assumed on which it can be the function of something to do something even when it fails to do that thing, and conversely something can do something even when it is not its function. For example, the function of the heart remains to pump blood even in a patient on life support whose heart can no longer fulfil this function. And conversely, while the heart in fact produces a rhythmic sound, that is not its function.

Although this is not the place for an in-depth review there are some advantages to taking the teleo-functionalist view that mental states are defined by the function of causally mediating between inputs and outputs. First, dry functionalism is not well suited to accommodating the possibility of abnormal psychology. It seems possible for there to be a 'wiring problem' such that a mental state fails to have its characterizing causal dispositions in relation to other states of its bearer. Malfunctioning systems seem to be possible. The teleo-functionalist can accommodate this possibility: it can remain the function of a system to have a certain causal profile when in fact it lacks that profile.

A converse complaint about dry functionalism is that it is too liberal. It seems possible for a complex system such as the economy of Bolivia, or some field of asteroids, to happen to enter an arrangement in which the system meets condition (F). It is implausible to think that these systems are therefore bearers of mental states. Teleo-functionalism does better: such systems are mere de facto causal unities: they have nothing with the *function* of causally mediating between inputs, outputs and other sub-systems of the system.

Probably the most impressive reason for favouring a teleo-functionalist approach to the mind is its congruence with the most developed naturalistic accounts of the representational content of mental states (Burge 2010; Neander 2017; Shea 2018; Schulte 2023). These accounts make use of a notion of function in order to solve problems for simpler appeals to causal role. For example, mere causal correlation or indication is neither sufficient nor necessary for genuine mental representation. It's not sufficient because a state can be correlated with all sorts of unnatural environmental conditions, not to mention causes which are more proximal than the object in the environment which is being targeted (retinal stimulations, mere object-surfaces). And it is not necessary because misrepresentation is possible: there are cases where a mental state is about a certain object or condition which is not present at all and *a fortiori*

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

cannot be among the causes of the representation. The theory that the object of a representation is that which it is the *function* of the representation to indicate promises to solve these problems. First, among the various causes, natural and unnatural, proximal and distal, the object stands out as that which the state functions to indicate. Second, in a case of misrepresentation when there is a failure of causal indication, the object can remain that which it is the function of the state to indicate.

These advantages support the general view that mental states are to be defined by the way they *should* interact causally with inputs, outputs, and other mental states of the same subject. Accordingly there is reason to transpose into a teleo-functionalist key the Shoemakerian approach to what it is to be a bearer of mental states, as follows:

> (TF) *x* is a bearer of mental states if and only if *x* has systems with the *function* of causally mediating in the right way between inputs of *x*, outputs of *x*, and other systems of *x*

A bearer of mental states is not a *de facto* causal unity but in this sense a teleo-functional unity.

The world of artefacts provides many examples of teleo-functional unities. A winebottle opener is a thing with parts (screw, lever) which have the function of enabling the thing to open a winebottle. An automobile is a thing with parts (engine, wheels) put together to move the whole thing around. The parts have a function in relation to a certain whole.

In order now to see how the approach promises to solve the problem of thinking parts, notice that there may be larger undetached parts of artefacts which are *de facto* causal unities as a side effect. For example, consider all of a car except for its passenger door. On the 'dry functionalist' view that an automobile is something with parts (engine, wheels) which in fact move it around, this 'sub-car' counts as an automobile: it too has parts (engine, wheels) which in fact move it around. It is 'internally' just like a car which has lost its passenger door. However moving the sub-car is not the function of those parts. The parts of the car were put together to move around the larger thing which also has a passenger door. So the teleo-functionalist theory of what it is to be an automobile (rightly) classes the car as the only automobile in the vicinity.

On the face of it, this is better than a 'minimalist' way of ruling out the three-doored undetached sub-car from counting as an automobile. One could insist that an automobile is really the smallest possible part of a car which is intrinsically like something with parts that could move it around. Perhaps this would be something composed of the engine, wheels, and drive shaft alone. Or more likely given typical over-engineering, something with only some thinner undetached core of

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

the metal of the drive shaft, some minimal inner band of rubber from each tyre, certainly no hub-caps. This minimalist view would explain why the sub-car is not an automobile. But it would do so at the cost of failing to classify paradigm automobiles as automobiles. The availability of the teleo-functionalist alternative makes such a revisionary view pointless.

Note that a teleo-functionalist proposal does not insist that the property of being an automobile is 'maximal': it does not insist no automobile could possibly have an automobile as a large proper part. It is consistent with teleo-functionalism there might be an automobile with another automobile as a large proper part. This is an advantage of the proposal because it is hard to see why there couldn't be such an automobile: consider a kind of James Bond car whose passenger seat has its own smaller engine and wheels, designed to detach as a whole for a quick getaway. This mini-car, unlike the accidental sub-car, does have parts with the *function* of moving it around. It is an automobile, which is a large part of another automobile. Maximality is at best a true generalization about actual automobiles.

We can now see how the teleo-functionalist approach to what it is to be a bearer of mental states promises to select among an animal and its various parts. The approach is congenial to the possibility of artificial psychology: there could be a robot with systems designed to causally coordinate it. But in the case of natural systems a notion of natural function rather than intentional design needs to be applied.[68] Although a range of theories are likely to deliver a similar result, consider the view that the natural function of a system is the activity whose adaptive success in its ancestors explains the current presence of the system. For example the heart in fact does many things, such as producing a rhythmic beat, but it is the circulation of blood which explains why there is a heart there now. So that identifies the function of the heart.

What is the function of the cognitive systems in a human animal? It is plausible that they function to coordinate the animal: it is the past success in coordinating animals in a complex environment which explains why the systems are there. They in fact causally coordinate various undetached parts of animals, but that is just a side effect of their evolved function. So it is the animal alone which meets condition (TF). The animal is the bearer. Other brain-overlapping objects only meet the condition of being an x with parts that function to coordinate some distinct y.

The approach also addresses a related puzzle which does not concern proper parts of animals. In Section 2, we found no reason to deny the general possibility

[68] See (Rubner 2022) for a review of naturalistic theories of proper function in the philosophy of mind.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

of coincident objects, discernible by their persistence conditions and other sort-proprietary properties. One might think that coincident with an animal is a particular sum or fusion of particles, something which is discernible from the animal because the animal coincides with different such fusions from time to time as it exchanges matter with its environment. A fusion which coincides with an animal at one time soon scatters into the environment. But when it does coincide it matches the animal in various intrinsic respects. We don't ordinarily count this thing as a bearer of mental states but is there any principled reason to suppose that it is not? The teleo-functionalist proposal gives an answer: it is not contribution to the persistence of these fusions (they persist as long as their particles exist, no matter how scattered) nor contribution to the reproduction of any 'ancestors' of such fusions that explains the presence of human cognitive systems. These coincident fusions may fleetingly count as de facto causal unities but they are not teleo-functional unities.

So an independently motivated naturalistic metaphysics of mind selects the animal from other candidates. On this way of looking at things, animals are not selected as bearers of mental states because that fits the way we ordinarily talk, or because animals are the only composite objects which exist in the vicinity, but because of the real nature of the mental.

If our thoughts and experiences are by their nature related in this way to teleo-functional systems, then – in a major difference from the minimal realizer approach – there is no evident reason to distinguish the bearer from the thing which is the charitable assignment of reference to our self-representations, viz. the animal. Thus stands the traditional simple view that we are the bearers of our own thoughts and experiences.

4.4 Dicephalus

The description 'two heads and one body' captures the appearance of dicephalus conjoined twins. Like other varieties of conjoined twinning the case results from incomplete division of the zygote at an early stage of development. Such cases are rare, and even more rarely do they survive to term. However there is one well-known case, Abigail and Brittany Hensel, in which dicephalus twins survived to lead a relatively normal human life.

Such a case forms the basis of the following argument against animalism:[69]

(1) In the Hensel case there are two of us.
(2) In the Hensel case there is one animal.
(3) So we are not animals.

[69] See (Campbell & McMahan 2010) for an anti-animalist argument along these lines.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

In support of premise (1) is the strong appearance of distinct agents and thinkers, each with their own face, each capable of independent speech, each in full control of one arm. Distinct names are given by the family, and distinct driving licenses are given by the law. These appearances correlate with the underlying fact that there are two independently functioning brains, each supporting a unified set of mental phenomena.

In support of premise (2) is the point that there exists just one large skin-bounded object. Although there is some duplication of organs below the head in the Hensel case (there are two hearts for example), the various parts of the large object in many respects work together as a harmonious metabolic whole. Of course it is not quite the same shape as a normal human animal (it has two heads) but it is in many respects similar to a single human animal.

If these premises are true then the case shows that there are more of us than there are animals. The animalist's plural identity claim is false. Although the fact that brains and animals are normally one-one correlated conceals this truth from us, the case of dicephalus brings to light that we are more accurately counted by brains than by organisms. The simple alternative is that we just are brains. This is in line with the minimal realizer approach to bearers of experience. But there may be more subtle alternatives. One option is that we are something like Lockean persons, coincident with brains without being numerically identical to them.[70]

Can the animalist resist the argument? For the moment let premise (2) be granted. Are there options for the animalist who agrees that there is only one animal on the scene?

Premise (1), which says that there are two of us, could be resisted in two different ways. One could deny that there are *two* of us: according to a *disunity response* the case involves just one of us, albeit with radically disunified consciousness and agency. The other way to deny the first premise is to deny that the twins are two of *us*: according to an *exclusionary response* the case involves two unified thinkers of some kind, but not our kind.

Let's consider the exclusionary response. Animalism is generally understood to be consistent with the possibility of non-human persons who are not animals. It is consistent with the possibility of angels, or androids, to say that *we* are animals. The exclusionary response to dicephalus takes this further: the class of broadly *human* persons includes non-animals. The term 'human person' can be compared to the term 'jade'. It turns out that the class of mineral samples occupying a certain aesthetic role in jewellery-making comprises two quite different kinds of mineral, jadeite and nephrite. What the twinning case brings

[70] McMahan (2002).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

out is that the things that occupy the psychological, social, and legal roles of human life include things of two different kinds. The vast majority are human animals but a small majority are of a different kind. What kind? Options for the exclusionary response include the following: the twins are undetached heads, undetached brains, undetached animal-halves, or perhaps two things which massively overlap, each with just one arm and one head but sharing all and only the lower non-duplicated parts.

One concern with the exclusionary response is that it makes the truth that we are animals more parochial than expected. The things which share my nature turn out not even to include all earthly human persons. But it is not clear that this is a serious objection. It is not a serious objection to the theory that these samples of jade are nephrite that it has the consequence that not all jade on earth is of the same nature. Why shouldn't the world turn out a posteriori to be more metaphysically varied than superficial similarities would lead us to expect?

Is there something politically repugnant about an exclusionary response? Would it do the twin-persons an injustice to exclude them from 'us'. Not really. The exclusionary response can acknowledge that the manifest commonality between twin persons and ourselves is more significant for moral and legal purposes than the difference in metaphysical kind. By the same token, the manifest commonality of jadeite and nephrite can be agreed to be more significant for aesthetic purposes than their underlying metaphysical difference. Metaphysics isn't everything.

A different problem for the exclusionary response to premise (1) is what might be called a *generalization threat*. The threat is that once it is conceded that in this case, the persons are not human animals, it is difficult to resist saying the same about *our* cases, the ordinary case. If they are undetached brains or undetached human heads, for example, then why aren't we?[71]

We have already seen that the teleo-functionalist approach gives a response to the concern about undetached animal part thinkers in our case. According to the approach, the function of the cognitive systems in an ordinary animal is to causally coordinate that very animal, so the animal rather than any brain-overlapping proper parts of the animal is the bearer of experience. But it is an interesting question whether this approach can support the aspect of the exclusionary response which claims that there are two *non-animal* thinkers in dicephalus. If it is the function of both cognitive systems to causally coordinate animals which contain them, then, assuming there is only one animal, this animal is the bearer of both sets of experiences no matter that they are not

[71] Olson (2014).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

unified in the usual ways. That is, if there is one animal, then the teleo-functionalist approach would appear to support a disunity response better than an exclusionary response.

This may be too quick.[72] Perhaps the two cognitive systems, through learning and adaptation, acquire a function which diverges from the usual evolutionary function of whole animal coordination: perhaps each brain acquires the function of causally coordinating one of the relatively independent heads, or animal-halves. Thus could the teleo-functionalist approach support the conclusion that the undetached heads or animal-halves each come to be a bearer of experience, vindicating the exclusionary response on behalf of the animalist. Although we are animals, they are non-animal thinkers.

For animalists who deny the existence of any composite objects except for whole organisms this exclusionary response will not be an option. There are no undetached animal heads or brains or halves. A disunity response looks inevitable. If there is a single organism then there is only a single candidate for bearing the mental phenomena in the vicinity although these happen to fall into two disunified groups.

But whether it is motivated by teleo-functionalism or eliminativism about non-animals, a disunity response in this case seems incredible. Sometimes we can be 'of two minds', when conflicted, confused, or dissociated. But it is hard to believe that the Hensel twins are really one thinker speaking through two mouths.

4.5 Two Animals?

So far we have granted premise (2) that there is one animal in dicephalus. One might take the contrary view that dicephalus is actually a case of two organisms fused together (Liao 2006; Snowdon 2014). What could support this two-animal response?

One *prima facie* motivation parallels our discussion of persistence conditions in Section 3. We might reason as follows: we are animals of some kind; two of us can exist conjoined in a dicephalus case; therefore two animals of our kind can exist conjoined in a dicephalus case. In effect one reason from (1), and the negation of (3), to the negation of (2). Dicephalus is an empirical discovery about what is possible for the kind of animals that we are.

As in the debate about persistence conditions, however, this kind of inversion of an anti-animalist argument is likely to look unprincipled without some concordant theoretical backing for its rejection of (2). What does biological theory say about how to count organisms at a time?

[72] Thanks to Ian Phillips for discussion here.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

There are broadly two approaches in the current literature: evolutionary and physiological (Godfrey-Smith 2012). For purposes of evolutionary prediction and explanation, it may make sense simply to count biological individuals by events of fertilization. Such a criterion would count dicephalus as involving a single individual and not two, since the 'twins' originated from a single event of fertilization. However by such a criterion, even *non-conjoined* monozygotic twins would count as parts of, or in some other sense manifestations of, a single biological individual. But in the non-conjoined case, it seems obvious that there are two human animals. There are two integrated, self-maintaining units of a certain kind. Animalism, to be plausible, must be claiming that we are units of this kind rather than evolutionary individuals.

More promising are accounts of animal individuation in terms of physiological unity. A traditional account appeals to metabolic unity: the parts of an organism work together to maintain its structure despite turnover of material, by making use of energy and other resources from the environment. An organism is 'a metabolic whole' (Dupre & O'Malley 2009: 2; Godfrey-Smith 2012: 25). Another account appeals to the immune system: an organism is a whole composed of biochemically interacting parts which are all tolerated by an immune system (Pradeu 2012).[73]

Unfortunately for the two-animal response, these physiological conceptions of organism unity appear all to count a single animal in the case of dicephalus.[74] There is a single fused whole whose parts are physiologically connected in the relevant ways. The left and right halves of this whole are mutually biochemically interacting and immunologically tolerated. There is a unified circulatory system. The parts generally function together.

However we should question, following our discussion of persistence in Section 3, whether a monolithic notion of 'metabolic unity' ought to be assumed across the biological world. The metabolic and immunological theories presuppose that organisms of all kinds are individuated in the same way. It is not clear that this philosophical presupposition is true to the facts. Godfrey-Smith has suggested that 'the evolution of animals produced a new kind of biological unit' (2016b: 776). 'Animals with nervous systems are unified and coherent entities in another way' (2016b: 794). His point is that the kind of integration and self-maintenance exhibited by animals includes sensing and moving as a coordinated whole. These distinctive capacities are realized by complex fast-signalling nervous systems.

[73] Clarke (2013) gives an account which combines physiological and evolutionary factors.
[74] Boyle (2020) presses this point.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

From this point of view, it seems unprincipled to overlook the fact that the Hensel twins have distinct central nervous systems. It is obvious that the Hensel twins, while often cooperative, exhibit two of the nervous system-based sensory-motor unities associated with the evolution of animals. Why should broadly 'metabolic' accounts of organism individuation ignore this aspect of the unity of animals? A more judicious approach would regard the unity of an animal at a time as complex, just as the persistence of an animal over time is arguably complex. It can sense and act as a unit *and* police an immunological boundary. The case of dicephalus is a case in which these aspects of animal unity diverge. A two-animal response can question whether in a case in which these aspects diverge, the aspects of unity which are shared with plants and single-celled organisms should dominate the aspects of unity which are distinctive of animals, and are especially complex in human animals. In any case, we should question the dualist habit of picturing psychology and biology as somehow exclusive domains. A biological animal is not a mere 'body' with a sort of computer which happens to be resting on top of it. An animal is structured around a nervous system.

Here is an analogy. A sovereign state has a central government which can cause its population to act in a coordinated way. It also has control of immigration and trade at a geographical border. Typically these forms of unity coincide, determining the same population and geographical extent. However, it is possible for immigration and trade control to cease between two sovereign states, *de facto* or *de jure*. In such a case we need not say that the two sovereign states have given way to one large sovereign state. If there are still two central governments then we can say that the two sovereign states are now parts of a single free-trade and free-movement area. Central governance plausibly dominates border control in the individuation of sovereign states.

In the same spirit one might take the view that dicephalus involves two animals. They might nevertheless compose a single thing which is unified in some of the ways that an animal is normally unified, as a free-trade area is unified in some of the ways that a sovereign state is normally unified. This thing is not itself an animal, but a rare kind of biological organism produced in an unusual way, through a failure of a process which usually produces two separated animals. We could call it a 'megazygote'. It need not be denied that the two animals in dicephalus are also organisms: they are things which are self-maintaining integrated biological individuals. So the suggestion is that the case involves two animals but three organisms. Would it follow that there are three thinkers? No, it remains plausible that the proper function of the human nervous system is to coordinate the animal which contains it, not an abnormal colony-like macro-zygote. So there are two animals and two thinkers.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

4.6 Split Brains

However a two-animal response is harder to believe in a structurally similar case.

Commissurotomy is a procedure which severs connections between the cerebral hemispheres of the brains. This has been performed on certain patients to stop focal seizures from spreading across the entire cerebrum. Patients who have recovered from these operations appear to behave normally most of the time. However special experimental conditions elicit behaviour which suggests a duality of consciousness. Information from the two sides of the visual field goes first to opposite hemispheres; movement of the right and left hands are likewise most directly caused by opposite hemispheres; and speech is only directly caused by the left hemisphere. This means that the division of the hemispheres can give rise to disunity in response to stimuli. If an object is kept in the left visual field only then the subject will verbally report that they see nothing. But if asked to draw what they see with a pencil in their left hand then they may well draw the object correctly. If the picture they have drawn with their left hand is then displayed in the right visual field and the subject asks why they drew it, they are likely to verbally confabulate. Another important finding is that these commissurotomy subjects are no better than chance at comparing objects as the same or different on either side of the midline of their visual field.

These effects are difficult to detect under normal conditions because normal side-to-side head and eye movement usually allows information about a single point to reach both hemispheres despite the lateralization of visual input. There are also various ways in which the hemispheres are liable to indirectly communicate, and to receive qualitatively similar information in virtue of being connected to a single body. But in the light of the underlying persisting neuroanatomical duality, it is quite plausible to conclude that these are merely epistemic difficulties in bringing to light what is a persisting duality of consciousness and agency.[75]

The duality may be less obvious than the duality in the case of dicephalus but the case might still be used as the basis of an argument that we are not animals.

(1) The split-brain case involves two of us
(2) The split-brain case involves one animal.
(3) Therefore we are not animals

Here a *two-animals response* of denying premise (2) looks less plausible than it did in the dicephalus case. Consider the human animal before the operation. Slicing through just a few centimetres of nerve fibres does not *seem* to be a way

[75] Nagel (1971) first brought commissurotomy to philosophical prominence. Schechter (2018) gives an outstanding philosophically-oriented review of the empirical literature.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

of creating two animals. Still, one might put such weight on the psychological aspects of animal unity, or more specifically human animal unity, that a two human animal verdict can be made to seem plausible. Still there is an awkward question, unparalleled for the two-animal view of dicephalus, about what happened to the human animal which was present before the commissurotomy. The procedure does not seem to be a way of destroying a human animal. Perhaps it is possible to respond that despite this partial disunification the original animal retains enough non-psychological unity to continue to exist. It is not destroyed. In which case, there are three animals on the scene, one persisting and two new. This is a strange view, which can be seem as simultaneously disputing (1) as well as (2). There are three of us after commissurotomy.

Likely much more attractive is to accept (2) but mount either an exclusionary or disunity response to premise (1).

First the exclusionary response according to which the two subjects or agents in the commissurotomy case are not among us. This is consistent with saying that *we* are animals. But then what are these non-animal beings? One option is to say that they are brain hemispheres. However this option suggests a generalization threat. If they are brain hemispheres then shouldn't we say that in the ordinary non-split case we are whole cerebra?

In important recent work, Schechter (2018) suggests a different exclusionary option. She takes the animalist view that in the ordinary case the thinker is a large animal with a whole cerebrum as a part. But in commissurotomy there are two thinkers which are each large animal-like things with just a single hemisphere as a part. Lefty, the 'right-hemisphere complement' contains all parts of the animal except for the right hemisphere. And then massively overlapping Lefty is Righty, the left-hemisphere complement, which contains all parts of the animal except for the left hemisphere. Attributing thought to these very animal-like non-animals at least does not so obviously suggest that we should attribute thought to some non-animal kind of thing in the ordinary un-split case.

Is this exclusionary option principled? Schechter points out that Lefty and Righty are each similar in form to a human animal who has had a single hemispherectomy. It is plausible that a single hemispherectomy animal would still be a thinker, although slightly smaller than before, and with some deficits. So it is plausible that Lefty and Righty are each thinkers, albeit ones with corresponding deficits.

However the problem of thinking parts should make us very wary of attributing mentality to a large undetached part of an animal on the basis of its similarity in form to an amputee animal thinker. On this basis, a multitude of undetached part thinkers would have to be recognized.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

There are also awkward questions about what would happen if commissurotomy were followed by single hemispherectomy. Pre-operation there is a single animal thinker, call it Whole. The commissurotomy then produces, according to Schechter, two large non-animal thinkers Lefty and Righty. But suppose next that the right hemisphere is removed. The right hemisphere is already disconnected from Lefty so presumably this leaves Lefty psychologically untouched. But now Whole has shrunk; Whole now coincides with Lefty. If, as Schechter plausibly claims, an animal can think after a hemispherectomy, then presumably it can also think when that hemispherectomy proceeded by way of a commissurotomy. So Whole and Lefty must be two coincident thinkers. This is an implausible consequence of Schechter's attempt to combine the claim that animals think in the ordinary case and the claim that non-animals think in the split case.

The minimal realizer approach to the thinking parts problem sidelines these problems. In no case do animals think. Hemispheres are the bearers of their thoughts. Whole cerebra are the bearers of ours.

What about the teleo-functionalist approach to the thinking parts problem? It suggests a disunity response to the argument from commissurotomy: there are not two thinkers of any kind but one disunified thinker. A teleo-functional unity can in fact fail to be unified in the way that it should. Moreover, large parts of a teleo-functional unity can be *de facto* unified without themselves being teleo-functional unities.

Here is an analogy to illustrate both points. Consider the following theory of what it is for some people to be a football team:

> (TEAM) Some people *xx* are a football team if and only if *xx* are supposed (by their intentions, contracts, management) to act together in certain coordinated ways.

Now consider some players *pp* who meet this condition when mid-match a thick fog descends down the long mid-line of the pitch making it very hard for the players on the left-hand side, call them *ll*, to communicate and interact with the players *rr* on the other side. Now in this circumstance, it may be possible for the smaller groups of players each to act together in ways mimicking a small football team, putting together attacking and defensive moves, in fact each better coordinated than the more inclusive *pp*. Still, the pluralities *ll* and *rr* are not really two football teams. They belong to one football team which has been caused to malfunction by the fog.

Now consider the right and left hemipsheres, or the two larger Schechterian undetached parts of the animal. Commissurotomy has caused each of these to become accidental causal unities, as the animal has become less unified. The animal is now something whose motor intentions sometimes fail to interact,

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

which cannot compare stimuli in the way that it should. The smaller entities don't fail to be integrated in these ways. Nevertheless, the surgical procedure does not alter the fact that it is the function of the cognitive systems to coordinate the animal. The surgery just causes some degree of malfunction. By (TF) then it is the animal and not the two undetached parts which should be counted as the single bearer of mental states.

This description of the case assumes a one-animal metaphysics of commissurotomy. As noted, it is just about possible that the right metaphysics of higher mammals weighs the cognitive aspects of unity so highly that the case should be regarded as a two-animal case. On the view that the function of human cognitive systems is animal coordination, (TF) would then deliver the result that there are two bearers of mental states. The 'failure' of intention integration across hemispheres would not actually count as a malfunction. Malfunctions will now derive from the abnormal overlap or attachment of two animals: for example, the kind of autonomous whole-body movement normally caused by the cognitive systems of a bearer of mental states will be to some extent disabled. Pre-operative autobiographical memories, which normally function to causally track one's own past, malfunction in the two new animals, causally tracking the past of the animal from which they have been split.

4.7 The Unity of Consciousness

Still, a one-animal reading looks more plausible than a two-animal reading. So (TF) is better seen as supporting a disunity response. For this reason, it will draw fire from theorists who insist on the necessity of the 'unity of consciousness' of states of a single bearer of conscious states. According to (TF) a bearer of mental states is something with systems with the function of integrating in various ways, but there is no necessity that it should always succeed. Some writers have held that the conscious states of a single subject of consciousness *must* be unified (Mackie 1976: 194; Bayne 2010). But it is very difficult to support such a strong theory. Of course whenever we introspect two experiences in a single act, we will, inevitably, find that the experiences are jointly accessible, and unified in that sense. But this is like the fact that whenever we see two pens at once they are inevitably both in a single field of view. It doesn't show that two pens cannot fail to be in a single field of view.

Does the difficulty we would have in projectively imagining in a single act of imagination the experience of a disunified subject count against the very possibility of such a subject (Schechter 2018: 46–49)? Well if there is no other explanation of our difficulty in imagining something, then that may well count as evidence of its impossibility. But in this case the fact that we are not

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

functionally and anatomically disunified in the same way seems to be a better explanation of the difficulty. Compare: our difficulty in projectively imagining what it is like to be a bat doesn't show that bats are not possible subjects of experience. Our difficulty in imagining what it is like to be a bat is better explained by our having a very different constitution to bats. Perhaps if simulating a subjective point of view was the only way of understanding others' mental states, then we should hesitate to countenance a disunified subject. But as our discussion of minimalism and teleo-functionalism show, theorizing about mental states can draw on more than just the subjective perspective.

Are there other reasons for thinking that subjective experience creates a special problem for the disunity response? It is plausible that any conscious experience in some sense contributes to 'what it is like' for the subject of that experience. With contribution spelt out in simple difference-making terms, the principle might be this: for any conscious experience of a subject, if the subject had lacked that experience, then there would have been a difference in what it is like to be that subject.[76] This principle does create a problem for the disunity response: suppose, as seems possible, the left and right hemispheres each realize experiences of exactly similar qualitative character. Suppose for *reductio* that these are the distinct experiences of a single animal. Then, for each experience, it fails to be true that if the subject had lacked that experience, then there would have been a difference in what it is like to be that subject. The other experience would have been enough for things to be qualitatively that way for the subject. This contradicts the principle. Therefore these are not distinct experiences of a single animal.

In response, one should question the principle. Why can't 'what it's like' sometimes be partly overdetermined? It is widely agreed that constitutive contribution or grounding of a fact can be overdetermined by facts which are individually sufficient. The fact that Alice is in the car is not barred from contributing to the fact that there is at least one person in the car by the fact that Bob is also in the car. Both facts contribute even though each would have been enough by itself.[77]

Another objection to a disunified bearer reading is that it cannot adequately describe the psychological situation. Suppose an intention to cooperate with the experiment is realized in the left hemisphere but no intention to cooperate with the experiment is realized in the right hemisphere. To simply say that the subject intends to cooperate will fail to describe the situation psychologically, and fail to predict and explain the behavioural dissociation. What needs to be added, one might think, is that there is also a subject who does *not* intend to cooperate.

[76] This is a construal of comments about commissurotomy in Peacocke (1983: 176–179).
[77] See (Snowdon 2014: 173–174) for related points.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

In response to this, it can be said that the animal intends (and bears mental phenomena generally) relative to brain parts. In the ordinary case, there is no need to make this relativization explicit. But when an animal has undergone commissurotomy, we may need to make this explicit. The animal intends relative to one brain part, but not the other.

Isn't this relativization response really conceding that brain parts are the real thinkers?[78] Not obviously. A community of one-handed people might never need to make explicit the relativisation of waving to a hand. These people simply describe themselves as waving or not. Perhaps it is only when they make the *a posteriori* discovery of a two-handed person who waves with one hand but does not wave with the other that it dawns on them that waving is, and always has been, something one does relative to a hand. It doesn't follow that hands are the things that really wave. People wave, with their hands. Or suppose that an animal's thorax is damaged so that when one lung deflates the other inflates and vice versa. An animal's lungs are usually coordinated so an animal can simply be described as breathing in or breathing out without qualification. The damaged animal however needs to be described as breathing in with one lung but not the other. It doesn't follow that lungs are really the things that breathe in and out. Animals breathe in and out, with their lungs.

In sum, a disunity response to split brains flows fairly naturally from a teleo-functionalist account of bearers. And reservations about the possibility of disunified consciousness are not easy to justify.

But we can end on a conciliatory note. Teleo-functionalist and minimal realizer accounts have been presented here as rivals. But it is an option to regard them in a more pluralistic spirit, as defining different ways in which mental phenomena are metaphysically dependent upon bearers. We don't have to choose. It is possible for something, by its very nature, to be related in one way to one particular thing and in another way to another particular thing. So in the case of commissurotomy we can say that there are two bearers of one kind, and one bearer of another.

This pluralist option would leave various questions outstanding for further research.

In Section 2, the relationalist view was canvassed that a conscious sensory experience involves a certain subject's awareness relation to objects, and therefore has a bearer. A question now is whether this bearer should be identified with the larger teleo-functional system, the minimal realizer of the experience, or neither. A naturalistic teleo-functional analysis of the awareness relation would

[78] Bayne (2010: 272) raises something like this objection.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

presumably favour the first option. But no such analysis has yet been developed. The awareness relation is more commonly treated as primitive.[79]

Finally, where would the pluralist option leave the question of what we are? On the face of it, recognizing different ways of being a bearer wouldn't significantly weaken the contention that animals are the epistemic best fit for our self-representations, and so that's what we are. An analogy: suppose we recognize a sense in which only a minimal part of the face of Tom the cat bears its grin. This wouldn't weaken the contention that the best fit for our Tom-representations is the whole cat.

[79] See for example (Brewer 2011: 94 and Pautz 2017).

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

References

Anscombe, G. E. M. (1975). The First Person. In Samuel D. Guttenplan (ed.), *Mind and Language*. Oxford University Press, 45–65.

Ayer, A. J. (1963). The Concept of a Person. In *The Concept of a Person and Other Essays*. London: Macmillan, 82–128.

Ayers, M. (1991). *Locke*, Vol. 2. London: Routledge.

Bailey, A. (2015). Animalism. *Philosophy Compass* 10 (12): 867–883.

Baker, L. R. (2000). *Persons and Bodies: A Constitution View*. New York: Cambridge University Press.

Bayne, T. (2010). *The Unity of Consciousness*. London: Oxford University Press.

Bennett, K. (2004). Spatio-temporal Coincidence and the Grounding Problem. *Philosophical Studies* 118 (3): 339–371.

Billon, A. (2015). Why Are We Certain that We Exist? *Philosophy and Phenomenological Research* 91(3): 723–759.

Block, N. (2011). Perceptual Consciousness Overflows Cognitive Access. *Trends in Cognitive Sciences* 15 (12): 567–575.

Boyle, A. (2020). Conjoined Twinning & Biological Individuation. *Philosophical Studies* 177 (8): 2395–2415.

Brewer, B. (1995). Bodily Awareness and the Self. In J. L. Bermudez, A. J. Marcel & N. M. Eilan eds., *The Body and the Self*. Cambridge, MA: MIT Press, 291–303.

Brewer, B. (2011). *Perception and Its Objects*. Oxford: Oxford University Press.

Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.

Campbell, J. (1999). Schizophrenia, the Space of Reasons, and Thinking as a Motor Process. *The Monist* 82 (4): 609–625.

Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

Campbell, K. (1990). *Abstract Particulars*. Oxford: Basil Blackwell.

Campbell, T. & McMahan, J. (2010). Animalism and the Varieties of Conjoined Twinning. *Theoretical Medicine and Bioethics* 31 (4): 285–301.

Chalmers, D. (2003). The Content and Epistemology of Phenomenal Belief. In Q. Smith & A. Jokic eds., *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press, 220–272.

Chisholm, R. M. (1969). On the Observability of the Self. *Philosophy and Phenomenological Research* 30: 7–21.

Clarke, E. (2013). The Multiple Realizability of Biological Individuals. *Journal of Philosophy* 110 (8): 413–435.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Dennett, D. C. (1992). The Self as a Center of Narrative Gravity. In F. S. Kessel, P. M. Cole, & D. L. Johnson eds., *Self and Consciousness: Multiple Perspectives*. New Jersey: Lawrence Erlbaum, 103–115.

Dickie, I. (2015). *Fixing Reference*. Oxford: Oxford University Press.

Dorr, C. & Rosen, G. (2002). Composition as a Fiction. In R. Gale ed., *The Blackwell Companion to Metaphysics*. Malden: Wiley- Blackwell, 151–174.

Ducasse, C. J. (1942). Moore's Refutation of Idealism. In P. A. Schilpp ed., *The Philosophy of G. E. Moore*. La Salle: Open Court, 232–233.

Duncan, M. (2019). The Self Shows Up in Experience. *Review of Philosophy and Psychology* 10 (2): 299–318.

Dupre, J. & O' Malley, M. (2009). Varieties of Living Things: Life at the Intersection of Lineage and Metabolism. *Philosophy and Theory in Biology* 1: e003.

Frege, G. (1997). Thought. In M. Beaney ed., *The Frege Reader*. Oxford: Blackwell, 325–345.

Gallagher, S. 2000. Philosophical Conceptions of the Self: Implications for Cognitive Science. *Trends in Cognitive Sciences* 4 (1): 14–21.

Garfield, J. L. (2022). *Losing Ourselves: Learning to Live without a Self*. Princeton, NJ: Princeton University Press.

Geddes, A. (2013). Think Twice, It's All Right: Animalism, Disunity and the Self. *Proceedings of the Aristotelian Society* 113 (3pt3): 371–380.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.

Godfrey-Smith, P. (2012). Darwinian Individuals. In F. Bouchard & P. Huneman eds., *From Groups to Individuals: Perspectives on Biological Associations and Emerging Individuality*. Cambridge: MIT Press, 17–36.

Godfrey-Smith, P. (2016a). Mind, Matter, and Metabolism. *Journal of Philosophy* 113: 481–506.

Godfrey-Smith, P. (2016b). Individuality, Subjectivity, and Minimal Cognition. *Biology and Philosophy* 31: 775–796.

Goff, P. (2017). *Consciousness and Fundamental Reality*. New York: Oxford University Press.

Hacker, P. M. S. (2007). *Human Nature: The Categorial Framework*. Oxford: Blackwell.

Haddock, A. (2019). 'I Am NN': A Reconstruction of Anscombe's 'The First Person'. *European Journal of Philosophy* 27 (4): 957–970.

Hakkarainen, J. (2012). Hume as a Trope Nominalist. *Canadian Journal of Philosophy* 42: 55–66.

Hawley, K. (2005). Fission, Fusion and Intrinsic Facts. *Philosophy and Phenomenological Research* 71: 602–621.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Heim, I. (2004). Lecture Notes on Indexicality. Ms., MIT. http://ocw.nur.ac.rw/NR/rdonlyres/Linguistics-and-Philosophy/24-954Fall-2004/02105696-3CFC-479B-9568-0A2089C79B9D/0/lec_notes_oindex.pdf

Heyes, C. (2018). *Cognitive Gadgets*. Cambridge, MA: Harvard University Press.

Hoffman, P. (2012). *Life's Ratchet*. New York: Basic Books.

Hohwy, J. & Michael, J. (2017). Why Should Any Body Have a Self? In A. J. T. Alsmith & Frédérique de Vignemont eds., *The Subject's Matter: Self-Consciousness and the Body*. Boston, MA: MIT Press, 363–391.

Hood, B. (2012). *The Self Illusion: Why There Is No 'You' Inside Your Head*. London: Constable & Robinson.

Ismael, J. & Pollock, J. L. (2006). So You Think You Exist? – In Defense of Nolipsism. In T. M. Crisp, M. Davidson, & D. Vander Laan eds., *Knowledge and Reality: Essays in Honor of Alvin Plantinga*. Dordrecht: Springer, 35–62.

Jékely, G., Godfrey-Smith, P., & Keijzer, F. (2021). Reafference and the Origin of the Self in Early Nervous System Evolution. *Philosophical Transactions of the Royal Society* B376: 1821. https://doi.org/10.1098/rstb.2019.0764.

Johnston, M. (1992). How to Speak of the Colours. *Philosophical Studies* 68 (3): 221–263.

Johnston, M. (2007). Human Beings Revisited: My Body Is Not an Animal. In D. Zimmerman ed., *Oxford Studies in Metaphysics*: Volume 3. Oxford: Clarendon Press, 33–74.

Johnston, M. (2016). Remnant Persons. In S. Blatti & P. Snowdon eds., *Animalism*. Oxford: Oxford University Press, 89–127.

Johnston, M. (2012). *Surviving Death*. Princeton, NJ: Princeton University Press.

Johnston, M. (2016). Personites, Maximality and Ontological Trash. *Philosophical Perspectives* 30 (1): 198–228.

Joyce, G. F., Deamer, D. W., & Fleischaker, G. (1994). *Origins of Life: The Central Concepts*. Boston, MA: Jones and Bartlett.

Kovacs, D. (2016). Self-Made People. *Mind* 125 (500): 1071–1099.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kripke, S. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge, MA: Harvard University Press.

Kurtsal, I. (2022). Self-Determination in Plenitude. *Erkenntnis* 87 (5): 2397–2418.

Lee, G. (2017). Selfless Experience. *Philosophical Perspectives* 31 (1): 207–243.

Letheby, C. & Gerrans, P. (2017). Self Unbound: Ego Dissolution in Psychedelic Experience. *Neuroscience of Consciousness* 3: 1–11.

Letheby, C. (2021). *Philosophy of Psychedelics*. Oxford: Oxford University Press.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Liao, M. (2006). The Organism View Defended. *The Monist* 89 (3): 334–350.
Lichtenberg, G. (2012). *Philosophical Writings*. Translated and annotated by S. Tester. Albany: State University of New York Press.
Locke, J. (1975). *An Essay Concerning Human Understanding*, ed. P. Nidditch. Oxford: Clarendon (2nd ed. published in 1694).
Mackie, J. L. (1976). *Problems from Locke*. Oxford: Clarendon Press.
Madden, R. (2016). Human Persistence. *Philosophers' Imprint* 16: 1–18.
Madden, R. (2019). How Can Thought Select between Coincident Material Things? In B. Brewer & J. Cumpa eds., *The Nature of Ordinary Objects*. New York: Cambridge University Press, 149–172.
Martin, M. G. F. (2002). The Transparency of Experience. *Mind and Language* 17 (4): 376–425.
Maturana, H. & Varela, F. (1980). *Autopoiesis and Cognition: The Realization of the Living.* Boston Studies in the Philosophy of Science, Vol. 43. Dordrecht: Reidel.
Maurin, A.-S. (2011). An Argument for the Existence of Tropes. *Erkenntnis* 74 (1): 69–79.
McClelland, T. (2017). Against Virtual Selves. *Erkenntnis* 84 (1): 21–40.
McDowell, J. (1997). Reductionism and the First Person. In J. Dancy ed., *Reading Parfit*. Oxford: Blackwell, 230–250.
McMahan, J. (2002). *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.
Metzinger, T. (2004). *Being No One: The Self-Model Theory of Subjectivity.* Cambridge: MIT Press.
Millière, R. (2020). The Varieties of Selflessness. *Philosophy and the Mind Sciences* 1 (1): 1–41.
Moore, G. E. (1903). The Refutation of Idealism. *Mind* 12 (48): 433–453.
Moore, G. E. (1955). Wittgenstein's Lectures in 1930–33. *Mind* 64 (253): 1–27.
Morgan, D. (2019). Accidentally about Me. *Mind* 128 (512): 1085–1115.
Nagel, T. (1971). Brain Bisection and the Unity of Consciousness. *Synthese* 22 (3–4): 396–413.
Nagel, T. (1974). What Is It Like to Be a Bat? *Philosophical Review* 83 (October): 435–450.
Neander, K. (2017). *A Mark of the Mental: A Defence of Informational Teleosemantics*. Cambridge, MA: MIT Press.
Noonan, H. (1998). Animalism versus Lockeanism: A Current Controversy. *Philosophical Quarterly* 48: 302–318.
Noonan, H. (2010). The Thinking Animal Problem and Personal Pronoun Revisionism. *Analysis* 70: 93–98.
Noonan, H. (2019). *Personal Identity*, 3rd ed. New York: Routledge.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Nozick, R. (1981). *Philosophical Explanations*. Cambridge: Harvard University Press.

Olson, E. (1997). *The Human Animal*. New York: Oxford University Press.

Olson, E. (2003). An Argument for Animalism. In R. Martin & J. Barresi eds., *Personal Identity.* Oxford: Blackwell, 318–334.

Olson, E. (2014). The Metaphysical Implications of Conjoined Twinning. *Southern Journal of Philosophy* 52 (S1): 24–40.

Olson, E. (2015). What Does It Mean to Say that We Are Animals? *Journal of Consciousness Studies* 22: 84–107.

Papineau, D. (2021). *The Metaphysics of Sensory Experience*. Oxford: Oxford University Press.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. (2012). We Are Not Human Beings. *Philosophy* 87: 5–28.

Pautz, A. (2017). The significance argument for the irreducibility of consciousness. *Philosophical Perspectives* 31 (1): 349–407.

Peacocke, C. (1983). *Sense and Content*. Oxford: Oxford University Press.

Perry, J. (1979). The Problem of the Essential Indexical. *Noûs* 13: 3–21.

Pradeu, T. (2012). *The Limits of the Self: Immunology and Biological Identity.* Oxford: Oxford University Press.

Recanati, F. (2012). *Mental Files*. Oxford: Oxford University Press.

Rubner, A. (2022). Normal-Proper Functions in the Philosophy of Mind. *Philosophy Compass* 17 (7): e12852. https://doi.org/10.1111/phc3.12852.

Schechter, E. (2018). *Self-Consciousness and 'Split' Brains*. Oxford: Oxford University Press.

Schlick, M. (1936). Meaning and Verification. *Philosophical Review* 45 (4): 339–369.

Schulte, P. (2023). *Mental Content*. Cambridge: Cambridge University Press.

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Shea, N. (2018). *Representation in Cognitive Science*. Oxford: Oxford University Press.

Shoemaker, S. (1963). *Self-Knowledge and Self-Identity.* Ithaca: Cornell University Press.

Shoemaker, S. (1968). Self-Reference and Self-Awareness. *Journal of Philosophy* 65: 555–567.

Shoemaker, S. (1984). Personal Identity: A Materialist Account. In Sydney Shoemaker & Richard Swinburne eds., *Personal Identity.* Oxford: Blackwell, 67–132.

Shoemaker, S. (1999). Self, Body, and Coincidence. *Aristotelian Society Supplementary Volume* 73: 287–306.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Shoemaker, S. (2004). Functionalism and Personal Identity – A Reply. *Noûs* 38(3): 525–533.

Shoemaker, S. (2008). Persons, Animals, and Identity.*Synthese* 162 (3): 313–324.

Sider, T. (2001). Criteria of Personal Identity and the Limits of Conceptual Analysis. *Philosophical Perspectives* 15: 189–209.

Simons, P. M. (2016). External Relations, Causal Coincidence, and Contingency. In A. Marmodoro & D. Yates eds., *The Metaphysics of Relations*. Oxford: Oxford University Press, 113–126.

Smart, J. J. C. (1959). Sensations and Brain Processes. *Philosophical Review* 68: 141–156.

Snowdon, P. F. (2014). *Persons, Animals, Ourselves*. New York: Oxford University Press.

Sober, E. (1985). Panglossian Functionalism and the Philosophy of Mind. *Synthese* 64 (August): 165–193.

Soteriou, M. (2013). *The Mind's Construction: The Ontology of Mind and Mental Action*. Oxford: Oxford University Press.

Strawson, G. (2015). Self-Intimation. *Phenomenology and the Cognitive Sciences* 14 (1) : 1–31.

Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London: Routledge.

Taylor, H. (2020). The Relation between Subjects and Their Conscious Experiences. *Philosophical Studies* 177 (11): 3493–3507.

Tye, M. (2003). *Consciousness and Persons: Unity and Identity*. Cambridge, MA: MIT Press.

van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination. *Adaptive Behavior* 14: 157–170.

van Inwagen, P. (1990). *Material Beings*. Ithaca, NY: Cornell University Press.

Velleman, D. (1996). Self to Self. *Philosophical Review* 105 (1): 39–76.

Watson, B. (1968). *The Complete Works of Chuang Tzu*. New York: Columbia University Press.

Wieland, J. W. & Betti, A. (2008). Relata-specific Relations: A Response to Vallicella. *Dialectica* 62 (4): 509–524.

Wiggins, D. (2001). *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.

Williams, B. (1970). The Self and the Future. *The Philosophical Review* 79 (2): 161–180.

Williamson, T. (2007). *The Philosophy of Philosophy*. Malden: Wiley-Blackwell.

Zahavi, D. (2005). *Subjectivity and Selfhood*. Cambridge, MA: MIT Press.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Cambridge Elements

Philosophy of Mind

Keith Frankish
The University of Sheffield

Keith Frankish is a philosopher specializing in philosophy of mind, philosophy of psychology, and philosophy of cognitive science. He is the author of *Mind and Supermind* (Cambridge University Press, 2004) and *Consciousness* (2005), and has also edited or coedited several collections of essays, including *The Cambridge Handbook of Cognitive Science* (Cambridge University Press, 2012), *The Cambridge Handbook of Artificial Intelligence* (Cambridge University Press, 2014) (both with William Ramsey), and *Illusionism as a Theory of Consciousness* (2017).

About the Series

This series provides concise, authoritative introductions to contemporary work in philosophy of mind, written by leading researchers and including both established and emerging topics. It provides an entry point to the primary literature and will be the standard resource for researchers, students, and anyone wanting a firm grounding in this fascinating field.

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press

Cambridge Elements

Philosophy of Mind

Elements in the Series

Attention and Mental Control
Carolyn Dicey Jennings

Biological Cognition
Bryce Huebner and Jay Schulkin

Embodied and Enactive Approaches to Cognition
Shaun Gallagher

Mental Content
Peter Schulte

Affective Bodily Awareness
Frédérique de Vignemont

The Computational Theory of Mind
Matteo Colombo and Gualtiero Piccinini

Memory and Remembering
Felipe De Brigard

Non-physicalist Theories of Consciousness
Hedda Hassel Mørch

Neurolaw
Gregg D. Caruso

Animal Minds
Marta Halina

Human Reasoning
David E. Over and Jonathan St B T Evans

Personal Identity and the Self
Rory Madden

A full series listing is available at: www.cambridge.org/EPMI

https://doi.org/10.1017/9781009570497 Published online by Cambridge University Press