Introduction

The focus of this volume is on the intricate relation between self-consciousness, memory, the body, and personal identity. As its title suggests, it considers the nature of self-identification, and also – with Wittgenstein – questions whether there is a self. Its most distinctive claim is that personal memory is central both to personal identity – as many have recognised – *and* to self-consciousness. It follows that one must address the questions of self-consciousness and personal identity together. The *question of self-consciousness* concerns what self-consciousness consists in; its relation to the ability to self-refer, expressed by means of "I"; and to self-knowledge, through such faculties as memory and bodily awareness. The *question of personal identity* concerns whether the criteria for personal identity are psychological, bodily, or a combination of these; is it memories and character traits, or bodily continuity, that makes me the same person as someone who lived twenty years ago? A novel contribution of this volume, I believe, is to draw these questions together, showing their unexpectedly intimate connection. In so doing, it rehabilitates a version of the memory criterion for personal identity, one that it is benignly and not viciously circular.

The preceding issues, and the distinctive claim concerning self-consciousness and personal identity, arise within what I call the volume's *core debate*. This debate is presented within a *framing debate*. The core debate, concerning memory, self-identification, and personal identity, draws attention to what may be termed *the epistemology of self-knowledge*: the ways we have of gaining knowledge of ourselves, of which memory is a central example. The framing debate concerns *"I" as a device of self-reference*. The monograph argues that understanding self-consciousness involves analysing both of these features, a strategy

that echoes Evans's two-part elucidation of self-consciousness in *The Varieties of Reference*:

> The essence of self-consciousness is self-reference ... thinking, by a subject of judgments, about himself, and hence, necessarily, about a subject of judgments. [But] our self-conscious thoughts about ourselves also rest upon various ways we have of gaining knowledge of ourselves as physical things.[1]

Like Evans, I contend that in order to understand "I" – and how it expresses self-consciousness – one must grasp its role both as a device of self-reference, and in eliciting the grounds of self-knowledge. Under the latter heading, the volume links the treatment of proprioception or bodily awareness with that of personal memory, drawing parallels between the kinds of knowledge to which these faculties give rise. The core debate takes up the inner four chapters, and the framing debate the outer three chapters.

The framing debate [Chapter 1]

The debate concerning self-consciousness and personal identity is framed by a more fundamental debate concerning the relation of self-consciousness and self-reference, discussion of which opens and closes the book. The debate arises from an assumption prominent in the Analytic tradition, which I formulate – and defend – as the *Analytic Principle*. It says that self-consciousness must be expressed by use of a self-referring device with the properties of the first person.[2]

But how essential is the use of "I" to self-consciousness? The scenario of a linguistic community that had no such uniform device of self-reference offers a helpful way into this question. Various historical personages have self-consciously self-referred using their own name: Julius Caesar, Henry James, and Charles de Gaulle. In de Gaulle's *Memoirs*, his third-personal use of "de Gaulle" becomes ever more insistent. The general was reported as saying that whenever he was faced by a particularly knotty problem, he would ask himself "What would *de Gaulle* do in this situation?" One might imagine a *community of name-users* whose

---

[1] Evans (1982), p. 213.

[2] If that assumption is false, the core debate that I am concerned with cannot get started; some considerations in favour of the assumption are advanced in the concluding chapter.

linguistic convention echoes the great general's habit: each speaker uses his or her own name self-consciously to self-refer. Is there any reason other than convenience why the self-referring use of non-indexicals, such as proper names and definite descriptions, could not generally replace "I"? My conclusion is that the discussion fails to show that "I" is ineliminable, since the name-user thought experiment rests on an implicit interpretation of the name-users as self-conscious. The scenario attempts to separate behaviour exhibiting self-consciousness from "I"-using behaviour, but its strategy is undermined by the way that use of a first-person device and the intention to self-refer remain stubbornly intertwined.

The core debate [chapters 2–5]

Chapters 2 and 3 move on to the core debate, the epistemology of self-knowledge, focusing first on the impossibility of mistaken self-identification in the case of personal memory; subsequent chapters extend the immunity to bodily awareness. This debate rests on a unifying concept, one that has acquired the cumbersome name of *immunity to error through misidentification*.

It can be illustrated by a homely example. I had a temporary job after leaving school working in the travel goods department of the now-defunct department store, Barkers of Kensington. I know this because I distinctly remember working in the travel goods department there, and there are friends who would corroborate my reminiscence. It may be said that memory is fallible. But one thing is clear: the personal memory-judgement "I sold suitcases in Barkers of Kensington" is *immune to error through misidentification*. That is, whenever a speaker claims that a judgement is based on personal memory, then although they might acknowledge the possibility of its being mistaken, since memory is fallible, it would never make sense for them to suppose that they were mistaken because of an error about *who* sold the suitcases. If, for whatever reason, I come to doubt that I had that job in the department store, it will make no sense for me to continue to maintain, using the same justification, that nonetheless *someone* used to sell suitcases; I cannot say "Well, I distinctly remember the suitcases being sold, but I'm not sure it was me who was doing the selling". When, in contrast, the claim is made on the basis of testimony – a friend apparently recalls seeing me behind the luggage counter in the store – then there is no identificational guarantee. If the friend later confessed that they really could not remember whether it was me

or someone else who had been behind the counter, it would not be senseless for me to ask "Evidently someone was selling suitcases, but was it myself?"

The crucial claim requires careful presentation; in particular, as we will see, it must be formulated in terms of the continuous-verb form characteristic of personal memory ("I remember selling them", as opposed to "I remember that I sold them"). It expresses a key feature of memory-based self-identification. The guarantee is that if I make the memory-claim "I sold suitcases in Barkers", and then subsequently come to doubt it, this could not be because I was mistaken about *who* sold suitcases. This guarantee is known in the literature as *immunity to error through misidentification*, with its ungainly initialisation *IEM*. (This is the only initialisation or acronym used in the book.) It is defended at length in the present volume, as a feature of memory-based self-ascription, and also bodily self-ascription. The immunity was first discussed by Wittgenstein in a well-known passage in *The Blue Book*, though he does not consider it in relation to either memory or the body.[3]

As some writers have acknowledged – though without mounting the strong defence of it found in the present volume – the immunity is essential to self-consciousness. Through the distinctive first-person epistemology of which it is the hallmark, it shows how one knows about oneself in ways in which one cannot know about others. What has not been recognised is the connection between immunity to error through misidentification and the long-standing *memory criterion for personal identity*. The connection is understood by considering the most influential critique of the memory criterion. The criterion says that persistence of memories constitutes personal identity. More precisely: X at time t is the same person as Y at time t+n if and only if Y remembers – in an appropriately personal or "inside" way – events witnessed by, and actions performed by, X. There are many problems with this criterion. But the most serious, arguably, is that as Bishop Butler and Thomas Reid suggested, it is viciously circular: memory presupposes personal identity and so cannot constitute it. As a result of their arguments, the criterion has faced the persistent worry that it is infected with circularity.

This volume offers a partial rehabilitation of the traditional memory criterion, arguing that the sufficiency condition – that remembering

[3] Wittgenstein (1969). He did not use the term "immunity to error through misidentification", which belongs to Shoemaker (1968).

events witnessed by, and actions performed by, X, in an appropriately personal or "inside" way, is sufficient for being identical with X – is benignly and not viciously circular. The qualified truth in the criterion is found in the phenomenon of IEM, as follows: memory is sufficient for personal identity in that when "I was F" is made on the basis of personal memory, my grounds for believing that someone was F are grounds for believing that it was myself who was.[4] The claim rests on a guarantee of correct identification that applies to *all* memory-claims, whether true or false-in-detail; that is, it applies to cases where one says "I thought I remembered o-ing, but now I do not think I can have".[5] It does not rest simply on the factive character of memory-claims – that "I remember o-ing" implies "I o-ed" – as Reid and Butler held.[6] When these facts are understood, the circularity objection is undermined. Indeed, I will argue, the objection cannot even be articulated.

The guarantee of correct identification is a genuine one, not an artefact of a way of describing the situation, and it constitutes the truth that proponents of the memory criterion were attempting to capture. We can therefore see that the debates over self-consciousness and personal identity are essentially connected, through the interdependent concepts of memory and personal identity. A vital and long-hidden connection between personal identity and self-consciousness is disclosed; a cloud of metaphysics of personal identity is condensed into a drop of grammar concerning the epistemology of self-consciousness.

In Chapter 2, the IEM phenomenon is defended from the challenge of q-memory, which says that one can have memories originating in the past of a distinct subject, and in Chapter 3 from the charge of vicious circularity. The critique of q-memory offered here constitutes an attack on Parfitian reductionism concerning personal identity. Parfit holds that there is a level of thought about oneself more fundamental than the first personal, and that one can have memories originating in the past of a distinct subject which would introduce errors in self-identification; my treatment shows that these claims are mistaken. Chapters 2 and 3

---

[4] More precisely: X can legitimately judge, on the basis of remembering or apparently remembering o-ing, that they o-ed *if and only if* either X = the person who o-ed, or else no one o-ed (as we will see, with appropriate provisos concerning apparent memory).

[5] Since, as Reid argued, we must regard memory as essentially reliable, few sincere claims are entirely false.

[6] The point about factivity, and the significance of the continuous-verb formulation of personal memory, are elaborated later.

integrate the self-consciousness and personal identity debates in the way we have just seen.

Strategy for a resolution [chapters 3–5]

We saw earlier how the Analytic Principle – which says that self-consciousness must be expressed by use of a self-referring device with the properties of the first person – underlies the argument of this volume. The Principle is interpreted as involving what I term a *conceptual holism* between self-reference and self-consciousness. The diagnosis of conceptual holisms is the volume's key philosophical strategy, which has application to the problems of self-consciousness and many other areas of philosophy. To assert a conceptual holism is to assert an equivalence and interdependence between the concepts concerned, such that neither is more basic than the other: that is, a definition or understanding of concept X makes essential reference to that of concept Y, and vice versa; there is a relation of mutual presupposition between them. One cannot acquire the first concept without acquiring the second, and one cannot manifest understanding of one without also manifesting understanding of the other. These claims about acquisition and understanding arise from the nature of the concepts themselves, and are not merely empirical. Conceptual holisms should be contrasted with pairs of concepts that, although connected, do not form a relation of mutual presupposition, for instance "photography" and "picture". A photograph is a kind of picture, but "picture" is a more basic concept which can be understood without understanding "photograph".

As well as self-consciousness and self-reference, other central examples in this volume of pairs of concepts that stand in a holistic relation are proprioception and bodily identity, and memory and personal identity. With this last pair of concepts we return to the core debate presented above: that of self-identification and the memory criterion for personal identity. When it is recognised that memory and personal identity constitute a holism of interdependent concepts, one sees how the memory criterion – or, rather, a qualified version of it – can be virtuously and not viciously circular. In explaining the conceptual holism of the body and bodily awareness, later chapters of the book show how one may transcend the traditional opposition between psychological and bodily criteria for personal identity.

Still under the heading of the epistemology of self-reference, chapters 4 and 5 consider *immunity to errors of self-identification in the case of bodily awareness or proprioception*. These chapters offer a rapprochement

between Analytic and Phenomenological approaches. Although bodily awareness now commands more attention within Analytic philosophy of mind, the issues surrounding it remain enigmatic to the wider philosophical community. And even where Analytic philosophers have looked to the insights of Husserl and Merleau-Ponty concerning the body and self-consciousness, this volume argues they have not done justice to them. My discussion draws on Phenomenological understanding of the body's significance for self-consciousness in order to reject a sharp bodily–psychological divide. *Materialism about the body* – the position that the body is a purely material and epistemically outer entity – is an almost universal assumption outside the Phenomenological tradition, and I offer a fundamental critique of it. This critique argues that Cassam's standpoint of *materialism concerning self-consciousness* arises from an over-literal interpretation of the insights of the Phenomenological tradition.[7]

Just as the memory criterion for personal identity was re-interpreted and defended, I argue, so an analogous criterion for bodily identity is shown to yield a *self-conscious knowledge account* of bodily identity that undermines materialism about the body. This account says that "my body" is the body of which I have self-conscious (proprioceptive) knowledge, and which I can move basically, in Danto's sense, that is, I just do it, and do not do it by doing something else. The account is a development of Locke's view that to experience a limb as mine – to feel it when it is touched, to be conscious of it as hot or cold and as having other "affections", to have sympathy and concern for it – is necessary and sufficient for it to be mine.

Chapters 3 and 5 also defend the volume's most ambitious claims, which locate the self-consciousness and personal identity debates in an opposition between the *unity of consciousness* and the *unitary self* that may be discerned in Kant. "Unity of consciousness" normally refers to the unity of adjacent thoughts or sensations in the stream of consciousness; as a treatment of personal identity, it refers to a bundle theory on the Humean model. Some remarks by Michael Ayers, however, suggest how the concept can be developed: "after Hume and Kant, both deeply influenced by Locke, it can hardly be considered beyond all question that a unitary self is prior, whether epistemologically or ontologically, to the unity of consciousness".[8] The question at issue is whether or not the unification of consciousness through memory – and also through

[7] Cassam (1997a).
[8] Ayers (1991), Volume II, p. 270.

proprioception – simply involves consciousness of a unitary substantial self. Ayers rightly suggests that Butler and Reid assume that it does when they say that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity.

These chapters tease out what is involved in this difficult discussion, arguing that Ayers does not sufficiently separate a neglected Kantian concept of the unity of consciousness from that of Hume and modern successors such as Parfit. One should not bracket Hume and Kant together, since Kant's account of the unity of consciousness does not – as Hume's does – imply the reduction of personal identity to a bundle of perceptions or experiences. My treatment is firmly Kantian, and refers to the unity of the remembering and remembered subject. Kant's advance on Hume, according to a unity of consciousness interpretation, consists in his recognition that the unity is *necessary* and not, as Hume thought, contingent. Building on this last insight, chapters. 3 and 5 outline a *non-reductionist unity of consciousness account*, distinct from both main current alternatives in the personal identity debate: a reductionist version of the allegedly circular memory criterion, and a bodily criterion. The unitary self, interpreted in materialist and immaterialist terms, is rejected.

The question at issue, to reiterate, is whether the unification of consciousness through memory – and also through proprioception – simply involves consciousness of a unitary substantial self. Proponents of the Butler-Reid objection to the memory criterion hold that it must, on pain of unacceptable circularity. But I argue that such circularity arises only on the assumption that one or the other of the concepts of memory and personal identity has epistemic priority. This assumption should be rejected; the concepts of personal identity and memory are interdependent, constituting a conceptual holism in the sense outlined earlier. The opposition that Kant's work generates, therefore, is one between traditional approaches to personal identity, whether these propose a psychological or bodily criterion, and an approach that recognises the close connection between personal identity and self-consciousness proposed in this volume. (In fact, both approaches – not just that of the unity of consciousness – have roots in Kant's work.)

The framing debate concluded [chapters 6 and 7]

Chapter 6 returns to the framing debate, and considers questions of self-identification and self-reference in light of discussion of the epistemology of self-knowledge in Chapters 2–5, and also the relation of

self-location and self-consciousness. Chapter 7 returns to the relation of self-consciousness and self-reference, and considers a different way in which "I"-use might be separated from self-consciousness, thus generating a further challenge to the Analytic Principle. This challenge arises from the possibility of self-conscious non-language-users [CUT, and is the topic of the book's concluding chapter.] Since the topic is a very substantial one, the chapter can only form an introduction to future work. Many would argue that chimps and other higher mammals exhibit at least a primitive kind of self-consciousness, which thus appears detachable from "I"-use. The mirror test, the subject of much psychological research, strikingly illustrates their self-recognitional capacity – a core self-conscious capacity, involving a concept of oneself as an individual. Once chimps are familiarised with mirrors, they recognise their own mirror-image. When a spot of paint is placed on the chimp's forehead, they notice this in the mirror, and rub it off; they begin to use mirrors to look at previously unseen parts of their body such as the inside of the mouth. From a careful analysis of the chimp's behaviour and its implications, I argue that this is self-consciousness only in a primitive sense derivative from the human one, and so the Analytic Principle is preserved.

Chapter 7 outlines the standpoint of philosophical humanism to which this volume is committed, which regards self-consciousness as essentially a human phenomenon, extended only in a primitive sense to certain higher primates such as chimps. The standpoint is humanistic because it rejects the scientistic focus on subpersonal and neural levels of description and explanation, and stresses the irreducibility of the personal level in philosophy of mind. I argue that it is consistent with the attribution of primitive self-consciousness to non-human non-language-users, and with the Analytic Principle, which is thereby preserved.

Real progress in the philosophy of self-consciousness is inevitably slow and painstaking – yet it is possible, as I hope this book shows. I believe that the problem of self-consciousness offers a more fundamental and fruitful starting point for philosophy of mind than the traditional concentration on the mind-body problem or mental-physical relation. A focus on "I" and self-consciousness promises more concrete conceptual advances. I do not claim to have offered a solution or final word, but have attempted to cover new ground and open up neglected avenues of investigation.

1
Self-Consciousness and Its Linguistic Expression

The Introduction raised the question of whether use of "I" is essential to self-consciousness, citing various historical personages who self-consciously self-referred using their own name: Julius Caesar, Henry James, Charles de Gaulle, and Andy Hamilton, among others. Could there be a *community of name-users* whose linguistic convention generalises this rather egoistic use, in which each speaker uses only their own name to self-refer? Is there any reason other than convenience why the self-referring use of non-indexicals, such as proper names and definite descriptions, could not generally replace "I"? What is the connection between "I" and self-consciousness? These and related questions form the topic of Chapter 1.

The chapter begins by outlining this volume's two-part elucidation of self-consciousness – in terms of self-reference and the epistemology of self-consciousness. It characterises *the Analytic Principle* that self-consciousness is a phenomenon expressed by use of a self-referring device with the properties of the first person. It defends a strong interpretation of the Principle, involving what I describe as the *conceptual holism of self-consciousness and self-reference*, that is, the kind of account of these concepts that is often described as "no-priority". (The principle is Analytic in being widely accepted in the Analytic Tradition, not in being true purely in virtue of meaning.)

The question of whether self-conscious self-reference has to be indexical self-reference, and its implications for the relation between thought and language, is then pursued. The chapter concludes that although a community of name-users based on the de Gaulle model – who lack an indexical term of self-reference – seems a coherent possibility, the presupposition that its members are self-conscious, and the need to understand name-use in terms of "I", means that the example proves

nothing. "I"-use cannot be analysed in terms of self-reference, independently of the grounds of self-knowledge. So we have to turn to the other part of the two-part account of self-consciousness offered by this volume, which concerns the epistemology of self-consciousness.

Like other writers, I have been using the term "self-consciousness" rather unself-consciously. But a fundamental caution should be entered. Almost all writers, in disputing its constitution, readily suppose that self-consciousness is a well-defined and unified concept – perhaps assuming, with Evans, that it has an "essence". This assumption should not go unexamined. "Self-consciousness", in the sense which this volume addresses, is not entirely a philosophical term of art that denotes a theoretical concept. The requirement, characteristic of Analytic philosophy, that philosophical speculation should be grounded in pre-philosophical thought, encourages us to explore its everyday meaning.[1] Pre-philosophical use of the term "self-conscious" implies self-recognition or self-awareness, usually of an unwelcome or embarrassing kind. Self-awareness is insight into oneself which involves understanding how one's actions appear to others. Hence the compelling Strawsonian requirement, that to be self-conscious is to be able to view oneself as a person among persons.[2]

In its philosophical sense, however, self-consciousness implies a network of interrelated capacities, involving intelligence, rationality, and intentional agency. How extensive these capacities must be, for a creature to be described as self-conscious, is a matter of judgement; self-conscious behaviour exists on a continuum, and attributions of self-consciousness do not have to be unqualified. As we will see later, intelligence as such is not sufficient: use of tools, for instance, such as a bird dropping nuts from a height in order to crack them open, does not of itself imply self-consciousness. Self-consciousness perhaps admits of degrees across species, as we will later see in connection with chimps. But it does not admit of degrees within species, one would not say that one person had more self-consciousness than another, except perhaps in the everyday sense of being more prone to embarrassment.[3]

---

[1] As for instance Soames (2003) suggests.

[2] Strawson (1959). This vital claim is considered in the final chapter.

[3] It would be wrong, for instance, to describe anorexics or autistics as less self-conscious than non-sufferers; moreover, their condition is pathological and so the implications for the self-recognition constraint on self-consciousness are limited. Together with the question of attribution of self-consciousness to non-humans, these issues are pursued in the final chapter.

1.1 A two-part elucidation of self-consciousness: self-reference and the grounds of self-knowledge

Analytic philosophers have generally characterised self-consciousness and subjectivity in terms of the first-person pronoun. Elizabeth Anscombe, for instance, writes that "when we speak of self-consciousness we ... mean something manifested by the use of 'I'", while Gareth Evans comments that

> The essence of self-consciousness is self-reference, that is to say, thinking, by a subject of judgments, about himself, and hence, necessarily, about a subject of judgments.

José Bermúdez holds that the ability to have "I"-thoughts is distinctive of self-consciousness.[4]

A key development in the evolution of the modern concept of self-consciousness was the first-personal turn of Descartes's *Meditations*. But a modern understanding of self-consciousness really begins with the "Transcendental Deduction" in Kant's *Critique of Pure Reason*.[5] As we will see in Chapter 3, Kant understands that in self-conscious thought, I think of myself *as subject* as object. An excellent illustration of the subject-object nature of self-consciousness is found in Merleau-Ponty's example of touching: "When I press my two hands together, it is not a matter of two sensations felt together as one perceives two objects placed side by side, but of an ambiguous set-up in which both hands can alternate the roles of 'touching' and 'being touched'".[6] But this is a particularly vivid example of an everyday, though hard to grasp, phenomenon.

This notion of reflexivity is largely absent from Descartes's treatment. In the twentieth century, Analytic philosophers came to regard self-consciousness as a phenomenon expressed by use of a self-referring device with the properties of the first person, notably indexicality. Hence what I call *the Analytic Principle*, which says that self-consciousness is a capacity manifested through use of a device like "I", or through behaviour which must be interpreted by using such a device.

Clearly it is not use of the English word "I" that is essential, on this view, but grasp of the first person as a grammatical category, whether expressed as pronoun or verbal inflection. The Analytic Principle involves two claims. It asserts the *conceptual claim* that self-consciousness must be

---

[4] Anscombe (1981b), p. 25; Evans (1982), p. 213; Bermúdez (1998), p. 295.
[5] Descartes (1996); Kant (1929).
[6] Merleau-Ponty (1962), p. 93.

understood as a phenomenon expressed by use of a self-referring device with the properties of the first person. It also implies the *methodological claim* that understanding of the first person is the route to understanding of self-consciousness. This first chapter is concerned to defend these claims against critics who fail fully to acknowledge their significance.

The Analytic Principle has the merit of anchoring an elusive phenomenon in concrete linguistic practice concerning "I", and is a guiding assumption of this monograph. But it must not be understood too narrowly, as writers such as Anscombe perhaps do. Understanding "I", and hence self-consciousness, does not merely involve grasping its role as a device of self-reference; it also involves grasping the grounds of self-knowledge. Or rather, in grasping its role as a device of self-reference, one must also understand the grounds of self-knowledge. Thus Evans comments that although the essence of self-consciousness is self-reference, "our self-conscious thoughts about ourselves also rest upon various ways we have of gaining knowledge of ourselves as physical things".[7] First-person thoughts involve thinking about oneself in a way in which one cannot think about anything or anyone else.

Self-conscious or first-person thought – thought in which the first-person figures as grammatical subject – therefore requires a two-part elucidation. The first part concerns the *self-reference principle:* that thoughts or utterances with "I" as the subject-term refer to the thinker or speaker. The second concerns *the epistemology of self-consciousness*, expressed in the following platitude:

> *Distinctness Principle*: To have "I"-thoughts is to think about oneself in a distinctive way in which one cannot think about anyone or anything else.

A common way of expressing this principle has been to say that "I"-thoughts involve self-conscious self-reference. That was Anscombe's view, and it captures the essential fact that thoughts of type (B) involve something more than thoughts of type (A):[8]

(A) Hamilton is about to be bitten by a poisonous snake (where I do not realise that I am Hamilton).
(B) I am about to be bitten by a poisonous snake.

---

[7] Evans (1982), p. 213. He stresses self-knowledge as physical things to counter a possible Cartesian bias – Cartesian in the textbook mentalistic sense – in the self-reference principle.
[8] Anscombe (1981b).

Thoughts (A) and (B) have the same person as object, but dramatically different behavioural consequences; only if the subject has thought (B) will they take avoiding action. Drawing a contrast between these types of thought focuses attention on the *action-determining role of "I"*, and shows that "I"-thoughts are necessary for intentional action in any full sense.[9]

However, the contrast between (A) and (B) thoughts does not bring out the full implications of distinctness. In one sense, subject (B) thinks about itself in a distinctive way in which it does not think of others; subject (A), in contrast, thinks about itself in the same kind of way as that in which it thinks of others. But a full understanding of the epistemology of self-consciousness involves an important subclass of "I"-thoughts, those that exhibit *immunity to error through misidentification*. This category was illustrated in the Introduction by the example of personal memory-judgements. This topic, and the epistemology of self-consciousness in general, are discussed in subsequent chapters; the present chapter addresses self-reference.

1.2 Self-consciousness and conceptual holisms

Analytic philosophers have been hostile to the idea of in breaking circles of concepts. Bermúdez, for instance, is concerned about how language-learners penetrate the circle of "I"-use and "I"-thoughts, as grasp of the first-person pronoun seems to require prior first-person knowledge: "Any theory that tries to elucidate the capacity to think first-person thoughts through linguistic mastery of the first-person pronoun will be circular, because the explanandum is part of the explanans…".[10] Reflexive self-reference by means of "I" presupposes the capacity to think thoughts with first-person contents, he believes; it therefore cannot explain that capacity. For Bermúdez, self-consciousness is paradoxical, since if the abilities which underlie self-conscious thought can be explained only in terms of each other, their acquisition becomes inexplicable: how could young children "bootstrap themselves into self-conscious thought"?[11] His worry seems to be this: what allows creatures to acquire

[9] However, Perry's *action-generating principle*, that an "I"-thought is one that issues in self-directed action, cannot elucidate "I" in terms more basic than the first person (Perry 1979, especially p. 5). The action in question, "self-directed action", must be characterised in terms of "I".

[10] Bermúdez (1998), p. 16.

[11] "Self-consciousness depends upon the capacity to take a first-person perspective (to think first-person thoughts), but we have no understanding of how entry into that first-person perspective is possible" (Bermúdez 1998, p. 123).

the complex of capacities associated with self-consciousness, including the ability to self-refer? His view seems to be that if creatures' experience of their own bodily selves and the world around them is entirely "third-personal", it becomes mysterious how they could develop the capacity self-consciously to self-refer. He thus postulates a fundamental or primitive level of self-conscious or "self-specifying" thought allegedly abstracted from the first person. Though he recognises that the human case is paradigm, Bermúdez holds that primitive self-consciousness is a genuine phenomenon defined in evolutionary terms, and not just a way of speaking.[12]

Bermúdez's concern arises from a general Analytic worry about circularity that probably originates with Quine's critique in "Two Dogmas of Empiricism" of the analytic-synthetic distinction as viciously circular.[13] Quine complains that the logical positivists distinguish "analytic" and "universally true", without properly explaining that distinction. He argues that accounts of the analytic-synthetic distinction have proved either unsatisfactory or circular, typically trading on notions such as meaning and synonymy that are as much in need of explanation as analyticity itself.[14] Indeed, he continues, these notions comprise a family of concepts – "meaning", "synonymy", "necessity", "self-contradictory", "semantic rule", and "is defined as" – whose members, notably meaning, are also suspect. Quine argues that "if we are to explain analyticity … with help of cognitive synonymy as undertaken [earlier]", then we need "an account of cognitive synonymy not presupposing analyticity". For Quine, it seems, the mere existence of an explanatory circle of concepts is dubious. One can explain analyticity by appeal to synonymy, only if synonymy is a legitimate notion; and Quine will grant its legitimacy only if we can give it a clear, non-circular definition.[15]

Grice and Strawson responded that even if analyticity, synonymy, and meaning are too closely related to be capable of independent analysis, it

---

[12] Evans (1982); Campbell (1994); Cassam (1997a); Bermúdez (1998). Baker (2000) would be another example.

[13] Quine, "Two Dogmas of Empiricism", in his (1953) – see for instance Keefe (2002), p. 277. In its very nature, however, Analytic philosophy constituted a reaction to the earlier holism of British Idealists such as F. H. Bradley.

[14] Quine (1953), ref.

[15] As discussed by Soames (2003), p. 357; Chapter 16 has a lucid resumé of the debate over the analytic-synthetic distinction. Creath (2004) denies that circularity is at the root of Quine's critique: "Quine's worry must lie, not with circularity, but with some defect of each of the terms in this sequence … his basic demand is for behaviour criteria" (p. 49). See also Hylton (2007), Chapter 3.

does not follow that they are all irredeemably obscure.[16] This response, in my view, is essentially correct; meaning and synonymy require only as much explanation as analyticity itself, for these concepts form a conceptual holism, whose constituents must be explained conjointly. Quine aimed to undermine the logical positivist programme by showing that analyticity cannot be anchored in more fundamental concepts. But rather than concluding that the concept of analyticity is incoherent, one should instead reject the commitment – arising from the early Analytic project of logical analysis – to anchoring in fundamental concepts. That is, one should acknowledge the necessity of conceptual holism.[17]

Conceptual holism provides a distinctive interpretation of the Analytic Principle that self-consciousness must be understood as a phenomenon expressed by use of a self-referring device with the properties of the first person. That Principle may appear to assert the primacy of language over thought, but the alternative proposed here is that it asserts a conceptual equivalence or holism of thought and language, according to which neither has primacy. It is impossible to do justice in a volume on self-consciousness to the general question of the relation of thought and language. But it is approached here via the central instance of the relation between "I"-use and self-consciousness. This chapter argues that self-consciousness and use of "I" are inter-defined, so that neither can be understood independently of the other. Self-consciousness and self-reference must be analysed jointly. This strong interpretation of the Analytic Principle asserts a *conceptual holism of "I" and self-consciousness*: an explanatory interdependence or relation of mutual presupposition between self-consciousness and "I"-use. A conceptual holism of "I" and self-consciousness asserts a non-primacy, or relation of equivalence, between self-consciousness and self-reference.

There have been exceptions to the Analytic hostility to conceptual holism. Alasdair MacIntyre argues that narrative, intelligibility of a life, and accountability are no more nor less fundamental than personal identity: "The [former] concepts presuppose the applicability of the concept of personal identity, just as it presupposes their applicability…. The

[16] Grice and Strawson (1956).

[17] One might grant the legitimacy of circular analyses, and thus allow that the present volume's stance on self-consciousness and self-reference is coherent, while denying that it is, in addition, plausible. Thus one could hold that self-reference presupposes self-consciousness but not vice versa; arguments for the latter claim are found in the final chapter.

relationship is one of mutual presupposition". And Strawson defends the idea of a web of concepts.[18] Rather than always proceeding in the direction of greater simplicity, he holds, analysis can also delineate an elaborate network of connected items: "[Perhaps] we could not fully elucidate the concept of knowledge without reference to the concept of sense perception; [or] explain all the features of the concept of sense perception without reference to the concept of knowledge".[19] However, Strawson discusses the phenomenon of conceptual holism only in the most general terms, and has residual concerns about vicious circularity, commenting that some circularity in analysis is unavoidable, provided the circle of concepts is large enough. In contrast I would argue that there need be nothing objectionable about a holism between two concepts only.

To return to Bermúdez's worry in the light of conceptual holism. Some may argue that his worry is: what allows creatures to acquire the complex of capacities associated with self-consciousness, including the ability to self-refer? His answer is that creatures have experience that is self-specifying. To put the matter another way, suppose creatures don't have self-specifying experience; their experience of their own bodily selves and the world around them is entirely "third-personal". It then becomes completely mysterious how they could develop the capacity self-consciously to self-refer. I suppose one could bite the bullet and state that self-conscious capacities simply emerge together at some point. But this does not seem very satisfying.

This naturalistic account is one that I wish to deny. The "mystery" is a pseudo-problem. Bermúdez mistakenly holds that one can separate self-consciousness from the first-person perspective – a separation contrary to the strong interpretation of the Analytic Principle defended here. If one proceeds on the basis of separation, the elucidation of self-consciousness becomes intractable. But there is no need to go down this path, since the holism between "I" and self-consciousness can be explained in terms of the simultaneous acquisition of capacities, or the acquisition of a single complex capacity. Since Bermúdez claims not to be opposed to "local holisms" or conceptual circles as such, it is hard to see why he rejects a holism in the present case. There is no problem in stating that self-conscious capacities simply emerge together at some

[18] MacIntyre (1984), pp. 217–8; Strawson (1992), especially pp. 19–20. From another tradition, Ricoeur describes a "healthy circle" between time and narrative (Ricoeur 1984, p. 3).

[19] Strawson (1992), especially pp. 19–20.

point. "Self-specifying" is non-explanatory; it can only mean "first-personal".

Compare the relatively uncontroversial local holism between personal memory and factual memory – two fundamental forms of memory distinguished in the next chapter. One could not conceive of someone whose knowledge of their own past is derived entirely from merely factual memory, since if they could not retain personal memories, how could they retain information that they were told? Indeed, it is hard to conceive how they could even understand testimony concerning their own past – that is, the personal memory-reports of others. Yet the fact that the two memory-capacities must emerge together does not mean that they are inexplicable, since there can be an explanation of their joint emergence.[20]

The notion of conceptual holism is still somewhat obscure, so further elucidation is necessary. To reiterate, a claim of conceptual holism between *a* and *b* says that a definition or understanding of *a* assumes an understanding of *b*, and vice versa. One cannot acquire one concept without acquiring the other, nor manifest understanding of one without manifesting understanding of the other. Conceptual holisms should be contrasted with presupposition that is one-way, for instance that between "photograph" and "picture". A photograph is a kind of picture, and "photograph" cannot be understood without an understanding of "picture"; but "picture" is a more basic concept, and can be understood without understanding "photograph". The claim about acquisition is not empirical or psychological, but arises from the nature of the concepts themselves. One might, perhaps, speak – confusingly – of empirical holisms between concepts, in the sense of merely contingent associations of ideas. An example of an empirical holism might be that between morality and religion, in that many people regard the two as inseparable. Only those committed to a certain interpretation of divine command ethics – one that regards God and goodness as inter-defined, such that it is inconceivable that God could command anything other than the good – would regard "good" and "commanded by God" as a *conceptual* holism.[21] My claim is that, given the nature of the concepts in question, to acquire one concept is necessarily to acquire the other.

---

[20] Further examples of conceptual holism central to self-consciousness are discussed in Chapter 5.

[21] Wittgenstein regarded this as the only coherent version of divine command theory. This example is discussed in connection with the Euthyphro paradox in Chapter 4.

To claim a holism between two concepts is not to assert their identity, but there is a continuum of cases. Some pairs of concepts – "I" and self-consciousness for instance – almost coincide in their sense; others, notably memory and personal identity, clearly do not. All involve mutually constitutive criteria. Conceptual holisms occupy a middle ground between analytically inter-defined concepts, and cases where one concept has clear epistemic priority.

A philosophically uninteresting holism is that between "monarch" and "subject". It rests on the kind of definition that a dictionary could provide: a monarch is a ruler of subjects, and subjects are those ruled by a monarch. Genuinely conceptual, as opposed to analytic, holisms, in contrast, have often been denied or misconstrued in some way, and have significant philosophical consequences. They connect concepts that are roughly equivalent in importance. In this volume and elsewhere I propose and analyse holisms between memory and personal identity; proprioception and bodily individuation; my body and objectivity concepts; and belief and assertion.[22] Other defensible holisms include: concept and object; intention and action; natural law and causation; the right and the good; art and the aesthetic.

The conceptual holism between assertion and belief means that one cannot define assertion in terms of a more basic, independent notion of belief, nor can one define belief in terms of a more basic, independent notion of disposition to assert. This holism features in a philosophical treatment of belief comprising the assertion thesis, which says that avowals of belief are assertions of the fact believed, and that one cannot merely report a present belief without also asserting it; an interpretation of Moore's Paradox; and other claims.[23] A holism from a completely different field is that of art and the aesthetic. In the interpretation that I propose, it says that a society whose members appreciate natural phenomena such as sunsets aesthetically – that is, regard them as beautiful – also has to produce or at least recognise artworks, with a lowercase "a" and thus including craft products. The art-aesthetic holism may well constitute an expanded circle including concepts such as "beauty" and "craft". This is also true of the holism presently under consideration, between self-reference and self-consciousness. A postulation of holism between a pair of concepts is consistent with further concepts joining the circle; the essential claim is that neither of the

[22] Hamilton (2000).
[23] Defended in Hamilton (2000).

original pair is more basic than the other, and that they cannot be understood independently.

An alternative to regarding the Analytic Principle as resting on a conceptual holism, is to hold that it is a priori that self-consciousness and the ability to self-refer correlate. On this view, however, it does not follow that they are inter-definable; rather, the capacities are mutually dependent and this is known a priori. It should also be stressed that conceptual holism is distinct from the more widely accepted holism of belief. The latter is expressed in Wittgenstein's pregnant thought that "What we believe is not a single proposition, it is a whole system of propositions. (Light dawns gradually over the whole.)" [24] His insight applies equally to the acquisition of a conceptual system. For language-learners, light dawns gradually over the system of beliefs or concepts that they are in the process of acquiring. But the web of belief or concepts is a holism of attribution, not one of meaning; like Davidson's holism of desire and belief, it concerns the attribution of mental states or capacities to subjects. My present concern is with the very different notion of conceptual interdependence: of belief and assertion, or self-consciousness and self-reference. In defending a holism of belief and desire, Davidson is not saying that one cannot define the concept of desire independently from that of belief and vice versa, nor would such a claim be plausible. However, there is this connection between conceptual holism and the web of concepts: conceptual holisms show how concepts form a network, which the subject acquires gradually, like light dawning slowly over the whole.

To reiterate, the question of how language-learners penetrate the circle in question is one that exercises critics of conceptual holism, such as Quine and Bermúdez. The latter's paradox, that self-consciousness is required to master the first-person pronoun, yet mastery of the pronoun is required for self-consciousness, is dissolved when one understands that much, and perhaps all, concept-acquisition rests on conceptual holism. Holism concerning concept-acquisition is quite compatible with the necessity of grasping concepts prior to those in the "I"-circle. For instance, children grasp proper names including their own before they grasp "I", and use their own name as a self-referring term – though they have problems with the first person until they do grasp "I". If basic concepts are introduced to language-users by example rather than verbal definition, the existence of holisms is rendered less problematic. The

[24] Wittgenstein (1969b), para. 141, discussed in Hamilton (2013), Chapter X.

concepts are then shown to be anchored in our practices and activities. Far from betraying a vicious circularity, conceptual holisms are an essential feature of language, which an analytic approach to philosophical understanding must recognise.

Quine's position is reasonable only where there is a genuine doubt concerning the validity of the distinction in question.[25] He would not, presumably, deny that monarch and subject constitute a benign holism, since that holism is confirmed by the dictionary. In other cases, Quinean scepticism is appropriate – for instance when expressed as anthropological scepticism which holds that the distinction between magical and practical in traditional societies is a product of Western ethnocentrism. To members of such societies, the claim continues, all behaviour, including so-called magical practices, is practical – for instance, it makes the tribe plant crops on time. One cannot justify the distinction through a conceptual holism involving other concepts which also assume that distinction – such as, perhaps, the religious and the secular, and so forth. But a distinction is not questionable simply because it involves a conceptual holism.

In contrast to anthropological scepticism about the practical and the magical, scepticism about the distinction between self-consciousness and non-self-consciousness (or mere consciousness) – or between "I"-users and non-"I"-users – is not plausible. One may disagree on where to draw the boundary between self-conscious and non-self-conscious creatures; indeed one may, with Bermúdez, reject a conceptual holism concerning self-consciousness and "I". However, if one was seeking to show that self-consciousness is a well-defined concept, it would be circular to attempt to do so with reference to an allegedly unproblematic concept of "I"-use.

My conclusion therefore is that claims of conceptual holism do not involve egregious logical error. However, one cannot lay down in general terms the boundary between benign conceptual holism and vicious circularity. Rosanna Keefe has attempted to distinguish acceptable from unacceptable circularities, through a proposal that contrasts vicious inferential circularity and benign analytic circularity.[26] On this proposal, "x is F if and only if S judges x to be F" is analytically but not inferentially circular, since in determining whether S judges x to be F, it is not presupposed whether a actually is F, hence contemporary

[25] Quine (1953), Section III.
[26] The contrast originates in Humberstone (1997).

accounts of response-dependent concepts are not unacceptably circular. Keefe finds the contrast between inferential and analytic circularity unconvincing, however, arguing that some non-inferential circularities are indeed vicious; her conclusion, that no general account of acceptable circularity can be given, seems eminently sensible.[27] Philosophical analyses involving circularity must be assessed on a case-by-case basis; there may be no general explanation of why the Butler-Reid objection to the memory criterion is mistaken, while Reid's objection to the track-record argument for the reliability of memory is correct.[28] The general defence of conceptual holism is developed further in Ch. 3. Here we return to the particular holism of self-consciousness and self-reference.

1.3 Ability holism and the circularity of the self-reference rule

The claim of conceptual holism should be separated from that of an *ability holism* concerning the same concepts. The former claim says that understanding of self-consciousness and self-reference is interdependent. The claim of ability holism says that self-conscious and self-referential *abilities* are interdependent. It therefore makes a two-part claim: that to be self-conscious, one has to be able to self-refer; and that to be able to self-refer, one has to be self-conscious – that is, only selves can self-refer. (In a sense, of course, a conceptual holism involves an ability holism, since understanding of concepts is manifested in the ability to use them.)

The first part of the ability holism – that to be self-conscious, one has to be able to self-refer – is illustrated by the way that *the self-reference rule* itself presupposes self-consciousness. This rule – that "I" stands for or designates the person it is used by – is often regarded as the "meaning-rule" for "I". But Anscombe rejects it as either insufficient for self-conscious self-reference, or unacceptably circular. Even when reformulated as "'I' is the word each one uses when she knowingly and intentionally refers to herself", it fails to exclude unwitting self-reference, she argues.[29] When Oedipus unwittingly self-refers in declaring "The slayer of Laius must be captured", or when John Perry follows a trail

[27] Keefe (2002), pp. 280–1.
[28] Reid's objection is defended in Hamilton (2003), where it is argued that reliabilists are mistaken in assuming that memory can be justified in terms of more basic and independent concepts.
[29] Anscombe (1981b), pp. 22–3.

of sugar around a supermarket, wondering who created it and unaware that he himself did, each can be said knowingly and intentionally to refer to someone, where that someone = himself.[30] "Himself" in these cases is what Anscombe describes as the direct reflexive, a species of the reflexive pronoun explicable without reference to "I". That is, one can say that "Oedipus refers to himself", where "himself" is the direct reflexive, without implying that he self-consciously self-refers. This kind of self-reference is not sufficient for self-consciousness.

In contrast, if one interprets "himself" as the indirect reflexive, the rule becomes circular, Anscombe maintains, since "himself" can be explained only in terms of "I". Anscombe's argument seems to be that any sufficient rule must say: "I" is the word which each one uses (a) to refer to some person (b) knowing and intending that person to be him- or herself – where "himself" is the indirect reflexive, such that the subject's knowledge must be expressed circularly in terms of "I", in other words, "That I am the person referred to". This formulation, however, is circular.

The relevance of her discussion to conceptual holism is that Anscombe assumes that the self-reference rule, as a meaning-rule, must yield a non-basic or introductory explanation of "I" – one which does not presuppose an understanding of self-consciousness, or of "I". Others have shared this assumption. Thus Lucy O'Brien argues that circularity occurs in the use rather than the specification of the rule, namely the knowledge that a subject has that they themselves are using a term when they are; self-conscious self-reference presupposes at least that I know that I am the producer of this very token of "I".[31]

The requirement of an introductory explanation that is non-holistic is dubious in general, however, and cannot be satisfied in this case. The self-reference rule presupposes a grasp of "I" because in order to understand the generality expressed by "each one" – as in "'I' is the word each one uses…" – one must have an ability to distinguish between oneself and others, which is expressed, at least incipiently, in an ability self-consciously to self-refer. This ability is part of the background of conceptual capacities presupposed by an understanding of the self-reference rule, a background which involves self-consciousness. For instance, "the term which each one uses" implies intentional use, and it is doubtful that intentions can be ascribed to creatures lacking the

[30] Perry (1993).

[31] O'Brien (1994), p. 280, developed in her 2007 book. Her suggestion is perhaps implicit in Peacocke (1983) and Nozick (1981). The analysis that she offers will be criticised later.

capacity self-consciously to self-refer – at least given plausible assumptions about the sensitivity of intention to the concepts that may be attributed to agents. Insofar as a creature has at least primitive intentions, therefore, it must also have at least primitive self-consciousness.

We now turn to the second part of the claim of ability holism: that self-reference is necessarily self-conscious self-reference. It is true in that if Oedipus never self-referred self-consciously, he would not be self-conscious. The cases of non-self-conscious self-reference we have considered so far have been by language-users who are also capable of self-conscious self-reference. The latter must be the fundamental mode of self-reference, for there could not be creatures who self-referred only non-self-consciously. Anscombe, in her example of the society of "A"-users, somewhat surprisingly seems to disagree. "A" is defined as a word which each one uses, knowingly and intentionally, to refer to him- or herself, but which does not express self-consciousness, she holds. When they move a body part, an "A"-user has to check by observation that it is part of his own body, rather than someone else's, that has moved:

> He uses the name "A", as does everyone else, to refer to himself. So he is conscious of himself. So he has self-consciousness.
>
> But when we speak of self-consciousness we don't mean that. We mean something manifested by the use of "I" as opposed to "A".[32]

Hence, Anscombe argues, comprehending use of a term that everyone uses only to speak of themselves will not guarantee self-conscious self-reference.[33]

However, her example is under-described and seems incoherent. Since the "A"-users have to check by observation whether it is a part of their body rather than someone else's which has moved, they must lack proprioception, but then as will be argued in Chapter 4 they cannot be said to act at all, and so cannot be said to self-refer. This conclusion applies also to Anscombe's scenario of machines with scanning devices. Machines are not self-conscious and cannot be said to act intentionally, so they cannot self-refer (unself-consciously). Only selves self-refer.[34]

---

[32] Anscombe (1981b), p. 25.

[33] Anscombe (1981b), pp. 23–5.

[34] Her comment that the "A"-user has to check by observation whether it is a part of his body which has moved suggests that the crucial feature which "I" possesses and "A" lacks, and which Anscombe is groping for, is immunity to error through misidentification or IEM, the concept defined in Chapter 2.

1.4 Self-conscious self-reference and indexicality: the name-user scenario

As we saw in the Introduction, the name-user scenario dispenses with *indexical self-referring terms*, presenting a language in which self-conscious creatures self-refer without using "I". The scenario helps to elucidate the connection between "I" and self-consciousness, for which *indexicality* – which may be defined as "systematic variation of a term's reference with context" – is often regarded as necessary. We now assess the objection that this is merely a superficial linguistic characterisation of indexicality, one that rests on a deeper notion of indexical thought.

I will begin by looking at Anscombe's account because, according to many writers, she neglects the status of "I" as an indexical or token-reflexive: a device of self-reference, one that *each one* uses to self-refer. She denies that the reason why "I" is not a proper name is the "triviality that we perhaps would not *call* a word a proper name if everyone had it and used it only to speak of himself".[35] It seems that she regards the generality of the first-person pronoun – the existence of a term usable by everyone – as a triviality. This feature cannot be completely trivial, however, since it undermines her claim that the self-reference rule simply states which object the speaker refers to when they use "I". The rule does more than this; it says also that the speaker *intends* to refer to him- or herself. In grasping the rule, the speaker at least implicitly understands that "I" has the property that its reference changes according to who uses it. The rule specifies how that changing referent is determined – that is, it gives a sense for "I".

It is as if Anscombe glosses the self-reference rule thus: "'I' is an expression by means of which it is possible for someone (knowingly and intentionally) to self-refer". This gloss is clearly inadequate, because while natural languages possess many expressions that can be used to self-refer, only the first person and its cognates are defined *exclusively* as devices of self-reference. A speaker who does not realise that "I" is such a device lacks understanding of it. The first-person is a self-referring term or verbal inflection which, unlike the speaker's own name, is typographically identical – the same word – for all speakers.[36] "I" is used

[35] Anscombe (1981a), p. 24.

[36] A can of worms underlies the characterisation of words and sentences as "typographically identical". Those who favour sentences over propositions, in philosophy of language generally face a dilemma of an inadequate typographical characterisation, versus a propositional one that they wish to avoid (as discussed in Hamilton (2013), Chapter X.

comprehendingly only if it is understood that it is such a device; it is privileged as a device of self-reference in that it effects self-reference unambiguously.

Definite descriptions and proper names, in contrast, may be used self-consciously to self-refer, but they are not devices of self-reference exclusively. Oedipus grasps the use of the definite description "the person who brought trouble to Thebes", without realising that on this occasion he is the one who satisfies it. Later, after discovering that he does – that he himself brought trouble to Thebes – he would be able to use that definite description in order to self-refer self-consciously. A description can function as a self-referring term, but not as a *general device* like "I", which effects self-reference unambiguously – that is, as an indexical term. Indexicality, in the case of the first-person, may therefore just amount to unambiguous self-reference.

An example of self-referring definite description is the present writer's use of "the present writer", in order to avoid the intrusive "I". As we saw in the Introduction, various historical personages have self-consciously self-referred using their own name: Julius Caesar, Henry James, and Charles de Gaulle. Name-use can suggest a habit of dissociation, as it did with de Gaulle when, as Free French leader, he arrived in French Equatorial Africa in 1940: "There were shouts of 'De Gaulle, De Gaulle'. I was taken aback … I realised then that General de Gaulle had become a living legend … . There was a person named De Gaulle who existed in other people's minds and was really a separate personality from myself. From that day on I would have to reckon with this man, this General de Gaulle … ".[37] Similarly, after he had become an early media star in 1919, T. E. Lawrence announced that "Colonel Lawrence still goes on; only I have stepped out of the way".[38] Certainly name-use can imply a sense of seeing one's self as others do. "De Gaulle holds this opinion", unlike "I hold this opinion", assumes group membership and conventions that go with it. So the name-user convention seems to have something that "I" does not.

---

[37] Jackson (2003), pp. 52–3. In his *Memoirs*, Jackson explains, this third-personal use of "de Gaulle" becomes ever more insistent, appearing 50 times in the first volume, and 20 times in just the last two chapters of the final volume.

[38] Asher (1999), p. 355. "Lawrence both helped to create and then tried to deny the myth … . 'You see how false the praise is … how little the reality compared with the legend: how much luck: how little merit'" (op. cit., p. 354). Asher provides an excellent account of how the legend arose.

In another sense, however, name-use is more primitive than "I"-use. It echoes children's less than fully self-conscious use of their own name before they acquire a grasp of "I" – a grasp that comes interestingly late in linguistic development. When Jacob says "Jacob wants a biscuit", and becomes upset when not given one, he clearly means "I want a biscuit". We interpret Jacob as self-consciously self-referring, because although he does not yet grasp the first person, we infer from his normal linguistic development thus far that he is on the verge of doing so. Comprehending use of "I" goes with a pattern of behaviour inviting ascription of intention and self-consciousness. Given the behaviour without the use, one may be justified in ascribing the ability self-consciously to self-refer; conversely, given the use without the behaviour, one may decline to ascribe that ability. Indeed, "Jacob wants a biscuit" may not really require *interpretation* any more than "I want a biscuit" does; what Jacob is saying is transparent. We assume that he knows that he is Jacob, that is, that he knows his own name – if that really is an assumption, since children usually know their own names.

Self-referring uses of descriptions and proper names are peripheral to general linguistic practice. De Gaulle mostly used "de Gaulle" for rhetorical effect, and not as a replacement for "I", which he continued to use. Our question is whether there is any reason other than convenience why such uses could not generally replace "I". To return to the *community of name-users* – self-conscious language-users who use their own name to self-refer, apparently without loss of expressive power compared to our community of "I"-users. Each speaker adopts an *alternative self-reference rule* for a term used knowingly and intentionally to self-refer: "Their own name is the word which each one uses to refer to him- or herself". Thus there is a uniform rule, but no uniform term, across persons. The scenario has more and less radical variants. The most conservative abolishes only "I" and its cognates; in the most radical, all personal pronouns and first-, second- and third-person singular and plural verbal inflections are replaced by a uniform "third-person". The speaker's name would not count as an indexical term, since its reference does not vary with the context of utterance, namely, who uses it. "De Gaulle" always refers to de Gaulle, but only de Gaulle himself can use it to self-refer self-consciously. So there would be self-conscious self-reference without indexicality – or at least without use of an indexical term.

The question of whether the name-user community has lost anything essential to the means of linguistic expression is hard to answer. Is lack of a uniform device of self-reference, one that exhibits unambiguous

reference, merely an inconvenience – or does it signal a more radical loss?[39] Major inconveniences are as follows:

(1) Problems that arise when one does not know the speaker's name. With "I" the audience can identify the referent immediately – there is no doubt that the speaker is self-referring. Ventriloquism provides only a minor qualification of the immediate, unambiguous self-reference of "I", which rests on the very general fact of nature – as Wittgenstein would put it – that we can tell by movements of the mouth, and the apparent origin of the sound, who is speaking. The noises made by speech very often announce its spatial origin – in the sense of which mouth the words are coming from, not the identity of that person as Hamilton or whoever. But ventriloquism gives rise to problems not essentially different from the comparatively unusual situation of having to ask, of a lone voice in a crowd, "Who said that?"

In the name-user community, in contrast, it would not be clear whether they were self-referring by means of their own name or referring to someone else. In a small name-user community where everyone knows everyone else this would not be a problem; perhaps when a stranger appears, they could be given the names "Stranger 1", "Stranger 2", and so on. In a community of a more typical modern size, however, one would have to learn thousands of self-referring names in order to gain an equivalent of the practical use of "I".

We now see that "I" has two functions, as follows: If I say "AJH would like more beans", others at the meal would not know who wanted beans unless they knew who was AJH, even if they knew who was speaking; with "I" they just have to know who spoke. For clarity, I would have to say "AJH – speaking – wants more soup", but then the use of the name becomes gratuitous. So either names do not fully replace personal pronouns, or they do so with a device that functions as a replacement without names being used at all – in any case, they would no longer be names. When Lara screams "(Want) ice cream!", she may be given it if the donor realises who is shouting. "(Wants) ice cream!" is not an "I"-thought; it precedes both mastery of names and use of "I". When Lara has learned her name, she may shout "Lara wants ice cream", and may receive it even if the donor does not realise who shouted; but she may not, if they do not know her name. When she learns to say "I – Lara – want ice cream" that is not two words doing the same job, but two devices complementing each other – to close both loopholes, as it were. (This may be another way to try to elucidate "the intention to self-refer".)

---

[39] This view is advocated by Coval (1969), pp. 40–1. It originates with Wittgenstein (1975), para. 58.

Analogously, a secretary has to record the committee's votes, and puts a tick by the list of those attending as they vote. When Baldy says, "I vote against the wind farm", the secretary, who does not know him, asks "Who *are* you?" (not, "Who were you referring to?"). The example shows that in many cases "I" works properly only if those who hear or read it know who the speaker is. If Baldy had voted by saying "Baldy votes against the wind farm", the secretary might have paused before marking the sheet to ask, "I take it you were referring to yourself?"[40] It is not totally certain that Baldy is casting a vote, or describing how someone else is voting.

Thus we can see that, arising from the two functions of "I", are two inconveniences of name-use. There is the ambiguity of two people with the same name, and the further uncertainty – given that the listener does not know their name – of whether the person who refers to "Baldy" is self- or other-referring. Only a close-knit society where everyone had a different name, and everyone knew everyone else's name, would overcome these problems.

(2) How names are learned. Where "I" only is abolished, "I am AJH" could become "Name is AJH", and "What is my name?" – on the rare occasions on which it is asked – could become "What is name?" But in the more radical scenario where all personal pronouns and verb inflections are dispensed with, it is hard to see how the knowledge "I am AJH" could be expressed, except perhaps in the form "The person uttering this very sentence is AJH". However, knowledge of one's own name may not have to exist in propositional as well as practical form. The name-users might simply cotton on to people's names, including their own, without understanding identity statements involving them – as young children do before they acquire the use of personal pronouns, and indeed in grasping these also.

If the speaker were to forget their own name, she could not say "I've forgotten my name", though perhaps they could say "This person here", pointing.[41] Of course it is very uncommon for people to forget their names. But it may seem that this is a less serious matter than forgetting the use of "I", for in forgetting my own name, I have lost an item of

[40] Roger Squires also offers the following: In a noisy climbers' dormitory late at night, someone shouted "Shut up!". Famous climbers, hard men, looked about in amusement, and asked "Who do you think you are?" After the reply, "Whillans!", there was complete silence. "Me" would not have had that effect. (Don Whillans was a legendary climber from the north of England.)

[41] Since "This person here" substitutes for "I" only when supplemented by a pointing gesture it is not a true indexical. Even "The present speaker" may be ambiguous.

empirical knowledge, rather than a species of linguistic competence – the capacity self-consciously to self-refer. Indeed, one cannot simply "forget" how to use "I", since it is not a name and one cannot forget it as one can forget a name. This appearance results from *the seemingly indissoluble connection between the first person and self-consciousness*. However, it is conceivable that a speaker might indeed simply forget the *word* "I". Brain damage has bizarre effects, and one could imagine a victim asking, "What's that word, you know, the one for 'This person here [pointing to himself]'?" or "…the word that people use to refer to themselves?" It might seem that since "themselves" is a cognate of "I", the example is not persuasive. But note that the speaker is simply meant to forget a word, not lose the associated concept or capacity; they are able to provide a paraphrase. So both in our present notation and in name-user notations, the speaker might forget the word without losing the linguistic competence involved in self-referring. In the community of name-users, the need for a new name in cases of amnesia would be more urgent, but that is the only difference.

1.5 Implications of the name-user scenario

The conclusions to be drawn from the name-user scenario remain elusive, however. We are assuming that the name-users are self-conscious, and so must express their self-consciousness through use of their own name – thus perhaps begging the question of whether it is adequately expressed in this way. That is, we are stipulating *a prior fact of the matter* about whether they are self-conscious, independent of the behaviour that they manifest – which assumes that self-consciousness has a non-behavioural essence.

The alternative would be to describe a behaviour and then decide whether those exhibiting it are self-conscious. However, the behaviour is designed to be of a complexity that invites the attribution of self-consciousness. And indeed the *self-reference rule for name-users* – which we stipulated that the name-users followed – assumes grasp of self-consciousness, just as the self-reference rule for "I" does. The name-user rule is: "Their own name is the word which each one uses to refer to him- or herself". In order to understand the generality expressed by "each one", one must have an ability to distinguish between oneself and others, which is expressed, at least incipiently, in an ability self-consciously to self-refer. So it seems to follow trivially that we will regard the name-user scenario as one involving no loss of expressive power.

The name-user scenario tried to separate grasp of the first-person device from the intention self-consciously to self-refer; the name-users have the intention, but no uniform device by which to implement it. The scenario constitutes an attempt to separate self-consciousness from "I"-using behaviour – yet the first-person device and the intention to self-refer remain stubbornly intertwined. It does seem that in order to *interpret* the utterances of the name-users in the thought-experiment, one must use terms involving "I" – that is, the scenario is intelligible only because name-using judgements are interpreted implicitly as "I"-judgements.

It is true that a device of self-reference is simply a means of fulfilling an intention to self-refer, which, like any such means, exploits linguistic conventions. But in the case of "I", the intention to self-refer seems inbuilt because one does not have to distinguish self-consciously self-referring uses of the term from other uses; there *are* no other comprehending uses. (The use does have to be comprehending, of course; someone who *says* "I " is not guaranteed to understand it.) We attribute to the child that uses its own name an intention to self-refer without a grasp of "I", only because such a grasp seems incipient. Thus the appearance that "I" has been eliminated without loss of expressive content is illusory.[42] It seems that, for self-conscious creatures, the "method of description" using "I" is unavoidable, and not merely one possible conceptual scheme among others.

One could argue that the name-user thought-experiment smuggles in indexicality, as each proper name can refer only to the person to whom it belongs, even though confusion would arise if two people had the same name (such as the present writer, and comedy writer Andy Hamilton). Thus all proper names exhibit the rule "refers only to the person to whom it belongs" – a rule that is the essence of "I", since it is more important than the word. But could one understand what that meant, without having grasped a uniform *term* such as "I" on which the replacement term was imperfectly modelled? This seems unlikely. And so it seems that the name-user thought-experiment does not vindicate Anscombe's view that it is a triviality that "I" is a word which everyone uses only to speak of him- or herself.

My account of the scenario offers a particular interpretation of Perry's claim that "I" is the essential indexical, one that denies the primacy of thought to which he subscribes. For many writers would draw the following moral, or diagnosis, from the name-user scenario: that while

[42] This criticism applies also to Wittgenstein's discussion (Wittgenstein 1975, para. 58).

self-consciousness does not require actual use of "I", it does require having "*I*"*-thoughts*. On this view, self-consciousness involves having indexical thoughts, but does not require any (uniform) indexical term by which they are expressed. Hence the objection raised earlier, that my account offers only a superficial linguistic characterisation of indexicality, and that one should appeal instead to a deeper notion of indexical thought.

I wish to resist the suggestion that there is any such deeper notion. Thought is more basic than language only in the anodyne sense that "I am hungry" and "De Gaulle is hungry", both uttered by de Gaulle, express the same thought. That thought may be termed an "I"-thought, and one could say "De Gaulle has 'I'-thoughts but expresses them by using 'de Gaulle' instead of 'I'". But the attempt to make thought more basic than language in any deeper sense will prove non-explanatory, and should be rejected in favour of a *conceptual holism of thought and language*. By a "deeper sense" of the basic nature of thought, I mean a *thesis of the primacy of thought*, which says that unexpressed thoughts are more fundamental than their linguistic expression, and that they appear clothed in different linguistic garb.

An extreme version of this thesis is the "language of thought" hypothesis, that thought and thinking are performed in a mental "language", a symbolic system physically realised in the thinker's brain. (Despite the paradoxical terms in which it is expressed, this is a thought-over-(natural)-language hypothesis.) In less radical form, the thesis is assumed by those who regard indexical utterances as expressing perspectival knowledge of oneself and one's environment that is essential for action and self-consciousness. Thus if I believe that the treasure is buried in Durham, I will not start digging unless I believe that it is buried here; a particular indexical term may be dispensable, it is argued, but indexicality in the sense of beliefs with perspectival content is essential to self-consciousness. This is the sense in which Perry holds that "I" is an "essential indexical". "John Perry is making a mess", framed in the manner of de Gaulle, cannot explain his own behaviour without the additional explanation "and I believe that I am John Perry".[43] Since "Andy Hamilton" can express a conception of someone who happens to be me, *or* a first-personal self-conception, "Andy Hamilton is making a mess", asserted by myself, explains my adjusting the sack of sugar in my trolley only if it expresses a first-personal belief.

[43] Perry (1979), pp. 4–5.

"I" is indeed an essential indexical, but not in the way that Perry suggests. He fails to recognise that perspectival or indexical thought cannot be understood independently of indexical terms such as "I", "here", "now", and "this" – that is, independently of indexical language. Indexicality is a feature of both thought and language simultaneously; there is a conceptual holism between them. The claim that indexicality is essential to self-consciousness is therefore one about the nature of self-conscious thought and linguistic expression understood together; thought and linguistic expression are inseparable. Elucidation of perspectival thought proceeds through elucidation of "I"-thoughts, demonstrative thought, and so on – which involves elucidation of the features of "I", "this", and "that", and so forth. This conceptual holist position rests on what Bermúdez calls the *Thought-Language Principle*, which says that the capacity to think a certain range of thoughts must be analysed through the capacity for their canonical linguistic expression.[44] But it rests also on the converse principle: that the capacity for linguistic expression must be analysed in terms of the range of thoughts that are canonically expressed.

Many critics of the Thought-Language Principle would argue that the capacity to self-refer consists in having a special kind of self-conception, which a creature can have without being capable of using "I" or similar linguistic device. On their view, therefore, the Analytic Principle is incorrect. They may concede that, as humans, we have to interpret such self-referring creatures by using "I" or an equivalent linguistic device. We also have to use "I" to describe the content of primitive self-conscious experience to capture its self-specifying character. But they conclude that this does not show that self-conscious thought and use of a first-person linguistic device such as "I" are mutually dependent, in that one must be capable of self-conscious thought in order to use a linguistic device such as "I", and must be capable of using a linguistic device such as "I" to have self-conscious thoughts.

According to the conceptual holist position advocated here, this attempt to give primacy to thought-over language is misconceived. The human case, where thoughts are articulated in language, is central to self-consciousness; as the concluding chapter argues, one might attribute primitive self-consciousness to non-language-using creatures, but it does not follow that while thought is essential for self-consciousness,

[44] Bermúdez (1998), p. 27, 13.

linguistic expression is incidental. There are indeed forms of thinking that do not involve language – music is properly described as thinking in sound, sculpture could be regarded as thinking in bronze, marble, or whatever.[45] But it would be mistaken to suppose that some basic prior thinking *explains* how these accomplishments involving language, sound, or physical materials are possible. Rather, the accomplishments are some of the *forms that thinking takes*. While some kind of expression – musical, sculptural or linguistic – is necessary for thought as such, linguistic expression is necessary for *propositional* thought, I would argue.

We were trying to understand self-consciousness by contrasting self-conscious and non-self-conscious self-reference. (Recall that this was the original characterisation of the Distinctness Principle.) But "I"-use seems to be unanalysable; indexicality must be understood in terms of "I", "now", and the demonstratives. Other forms of self-reference are either accidental self-reference by creatures capable of self-conscious self-reference, or else rest on naming conventions that are themselves parasitic on "I"-use. Not achieving much success in understanding self-consciousness through direct examination of self-reference, one alternative approach might be to define self-conscious self-reference as the mode of self-reference by self-conscious creatures. As I have argued, there are no other genuine self-referrers; only selves self-refer. That is, all self-referrers are self-conscious self-referrers who on occasion self-refer accidentally or non-self-consciously. Some implications of this approach are considered in the final chapter of this book.

The main part of the volume, however, turns to the grounds of self-knowledge, the second part of the two-part elucidation of self-consciousness. To reiterate the claims made at the start of this chapter: self-conscious thought is about oneself, and involves thinking about oneself in a distinctive way in which one cannot think about anything else. In order to understand self-consciousness, we must not only understand self-reference, but also the epistemology of self-consciousness expressed in the Distinctness Principle: to have "I"-thoughts is to think about oneself in a distinctive way in which one cannot think about

[45] This conception of music is defended in Hamilton (2007a). It is opposed by the highly implausible Croce-Collingwood theory that art is mental intuition, only contingently expressed in physical form (Collingwood 1938), and, more plausibly, by the view that Beethoven's sonatas, Hepworth's bronzes, or Chippendale's chairs are not their thoughts, even though thinking, in the sense of problem-solving and experimenting, was required for their production.

anyone or anything else. In the following chapters, the implications of this principle are explored in depth. We will now see that – paradoxically, it may seem – self-consciousness involves *not* having to identify oneself as a subject in a range of central cases.

2

Memory and Self-Consciousness (1): Immunity to Error through Misidentification and the Critique of Quasi-Memory

As argued in Chapter 1, understanding "I" involves grasping both its role as a device of self-reference and the grounds of self-knowledge. When I think of myself self-consciously, I do so in a way that contrasts with thinking about myself non-self-consciously – as the examples of (A)-thoughts showed. However, there is a sub-category of "I"-thoughts that involves a distinctive way of thinking about myself, one in which I cannot think of others – a way involving distinctively self-conscious forms of self-knowledge, hence the heading "epistemology of self-consciousness" to label this issue. The kind of "I"-thoughts in question are those which are *immune to error through misidentification* (IEM), and they exhibit the Distinctness Principle: that to have "I"-thoughts is to think about oneself in a distinctive way in which one cannot think about anyone or anything else in a direct way. Wittgenstein was the first to recognise the category of IEM, but Shoemaker coined the phrase "immunity to error through misidentification" and articulated the phenomenon, while Evans gave the most developed account of it to date.[1] Other "I"-thoughts, in contrast, involve thinking of oneself in the same way as one thinks of others.

We encountered this category of "I"-thought in the Introduction. IEM is exemplified in personal memory-judgements such as the one presented there: "I sold suitcases in Barkers of Kensington." To reiterate, if the speaker presents the judgement as one based on personal memory,

[1] Wittgenstein (1969a); Shoemaker (1968); Evans (1982), Chapter 7.

then although they might acknowledge the possibility of its being mistaken, since memory is fallible, it could not make sense for them to suppose that they were mistaken because of an error about who sold the suitcases. If, for whatever reason, I come to doubt that I had that job in the department store, it will make no sense for me to continue to maintain, using the same justification, that nonetheless *someone* used to sell suitcases. I cannot say, "Well, I distinctly remember the suitcases being sold, but I'm not sure it was me who was doing the selling." When, in contrast, the claim is made on the basis of testimony – a friend apparently recalls seeing me behind the luggage counter in the store – then there is no identificational guarantee. If the friend later confessed that they really could not remember whether it was I or someone else who had been behind the counter, it would not be senseless for me to ask, "Evidently someone was selling suitcases, but was it myself?"

The present chapter offers a careful definition of IEM of memory-judgements, and a partial defence of a direct knowledge account of personal memory with which, I believe, IEM is associated. A preliminary discussion of the significance of IEM for self-consciousness is presented, and then objections to memory-based IEM are considered: those arising from information-garbling, from q-memory in both physicalist or functionalist terms, or in less specific formulations. It is argued that q-memory as normally presented rests on an unacknowledged assumption of an impersonal concept of information. This impersonal concept is criticised, and a personal concept of information defended in its place.

2.1 The epistemology of self-consciousness: "I"-as-subject and autobiographical memory

A vivid illustration of the kind of senselessness that IEM excludes is given by William James in his *Principles of Psychology*. In his discussion of self-consciousness, there is a curious footnote on the feeling of absence of self, which discusses a character called "Baldy":

> In half-stunned states self-consciousness may lapse. A friend writes me: "We were driving back from in a wagonette. The door flew open and X, alias 'Baldy', fell out on the road. We pulled up at once, and then he said 'Did anybody fall out?' or 'Who fell out?' – I don't exactly remember the words. When told that Baldy fell out, he said 'Did Baldy fall out? Poor Baldy!'"[2]

[2] James (1950), vol. I, p. 273n; the example is discussed by Anscombe (1981b).

The case is pertinent to us if Baldy's question was "Who fell out?" In such a situation, one may be completely disoriented, and not really able to say what has happened; perhaps Baldy's utterance is merely an expression of this disorientation. Where James goes wrong is in suggesting that the example shows how human consciousness need not imply self-consciousness. The kind of "lapse of self-consciousness" that James suggests is not possible: "Baldy" cannot be disoriented in such a way as to understand what has happened, but not realise who it has happened to. He cannot seriously be asking "Who fell out?"[3]

I wrote earlier that self-consciousness involves self-recognition. However, the phenomenon of IEM qualifies this truism, since there is strictly speaking no self-recognition in IEM cases. This point is made in the first clear description of the phenomenon, by Wittgenstein in *The Blue Book*, which contrasts "*I*"*-as-object* and "*I*"*-as-subject uses*:

> There are two different cases in the use of the word "I" (or "my") which I might call "the use as object" and "the use as subject". Examples of the first kind of use are these: "My arm is broken", "I have grown six inches", "I have a bump on my forehead", "The wind blows my hair about". Examples of the second kind are: "I see so-and-so", "I hear so-and-so", "I try to lift my arm", "I think it will rain", "I have toothache". One can point to the difference between these categories by saying: The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error, or as I should rather put it: The possibility of an error has been provided for ... On the other hand, there is no question of recognising a person when I say "I have toothache".[4]

This second category of "I"-judgements allows no possibility of error due to misrecognition; the judgement is immune to error through misidentification of the subject. Wittgenstein concludes that such

---

[3] Equally unintelligible is Mrs Gradgrind's death-bed pronouncement in Dickens's *Hard Times*: "'Are you in pain, dear mother?' [asked Louisa]. 'I think there's a pain somewhere in the room', said Mrs Gradgrind, 'but I couldn't positively say that I have got it.' After this strange speech, she lay silent for some time." Since "I am in pain" is an avowal – and thus completely immune to error – it is certainly IEM.

[4] Wittgenstein (1969a), pp. 66–7. "Immunity to error through misidentification" is a more accurate term than "criterionless self-ascription", which has also been used.

cases do not involve recognising a person, that is, they do not involve self-recognition.[5]

In citing "I have toothache" as an example of this category of uses, Wittgenstein invites misinterpretation of the phenomenon of IEM. "I have toothache" exhibits immunity to error through misidentification, but on a Wittgensteinian view, it is also an *avowal*, that is, it exhibits complete immunity to error. Avowals are exclusively psychological and almost exclusively present-tense utterances, and thus constitute a subcategory of "I"-as-subject or IEM uses. Examples include "I am in pain", "I believe that p", and "I intend to o". My primary concern here is not with avowals, however, but with the broader category of "I"-as-subject uses.[6]

This broader category also comprises bodily self-ascription, and ascription over time in the form of first-person memory-judgements. These two categories are my primary focus. One must separate the complete immunity to error exhibited by avowals – the authority of avowals, or what is often known as first-person authority – from the more restricted immunity to error which is IEM.[7] (Many philosophers will deny the authority of avowals. It is not my aim here to defend it, which I have done elsewhere.)[8] But in any case the debate is made clearer if such contentious examples are not selected as paradigms of IEM, in the way that Wittgenstein seems to.)

The idea that self-consciousness involves thinking about oneself in a distinctive way has been emphasised by Frege and by Perry. These writers nonetheless failed to recognise that such thinking involves immunity to error through misidentification. For Frege, the self-conception was a private one: I am presented to myself "in a particular and primitive way" in which I am presented to no one else.[9] He was not aware of the distinction between IEM and non-IEM cases, and so did not

---

[5] Underlying this conclusion is his principle that knowledge implies the logical possibility of doubt, which also underlies his controversial view that it cannot be said of me that I know I am in pain. In each case, knowledge is regarded as an active cognitive achievement. (These views are discussed in Hamilton (forthcoming, 2013, Chapter X). In the case of IEM and self-recognition, I believe that Wittgenstein's conclusion is highly plausible (see Chapter 6).

[6] The Wittgensteinian claim that avowals are immune to error of any kind is defended in Hamilton (2000) and (2008).

[7] Anscombe (1981b) makes this conflation, which is still found in the literature.

[8] Hamilton (2000) and (2007a).

[9] Perry (1979); Frege (1967), pp. 25–6. Frege's claim is discussed further in Chapter 6.

realise that the most distinctive mode of self-presentation is exhibited by the former. Some recent writers, notably Evans and Bermúdez, have acknowledged IEM's centrality to self-consciousness; the latter indeed has argued for a close connection between IEM and the essence of first-person judgements.[10] However, they do not defend IEM as a priori in the way it is defended here, that is, as constituting a necessary connection between the remembered and remembering subject, or between the limb-owner and the limb owned.

These may seem rather specialised aspects of the epistemology of self-knowledge, and their implications are not easy to grasp on first acquaintance. They are, however, crucial to the understanding of the "I"-as-subject use, and hence of self-consciousness. The present chapter, and Chapters 3 and 5, constitute an extended defence of IEM in the case both of memory and proprioception. This is the spine of this book's argument. To reiterate, the significance of IEM for self-consciousness arises, in the first instance, because first-person thought involves the Distinctness Principle: thinking about oneself in a distinctive way, as subject, a way in which one cannot think of anyone or anything else. This is shown by the fact that first-person IEM thoughts involve ways of knowing only of oneself, if of anyone. Non-IEM thoughts, in contrast, involve ways of knowing about oneself which are also ways of knowing about other people. The way in which I know that I weigh 12 stones, or – where the judgement is made on the basis of visual perception – that my hair is blowing in the wind, are also ways in which I can know that someone else weighs 12 stones, or that someone else's hair is blowing in the wind. Hence the possibility of misidentification in non-IEM cases; these are not "as-subject" uses.

The preceding considerations, I believe, show that "I" should be characterised as a self-referential indexical whose central uses exhibit IEM. How the guarantee of IEM arises is a key question. The strong conception of IEM as a priori, defended here, does not have a uniform explanation across the different cases. In present-tense cases, immunity to error in respect of misidentification is grounded in the complete immunity to error of avowals. (Recall that, in the case of avowals, no error at all is possible, while IEM-judgements show only partial immunity to error.) The appearance of a guarantee of "correct identification" in present-tense cases – which are mostly avowals such as "I have toothache" – arises

[10] Evans (1982), Chapter 7; Bermúdez (1998), p. 144 and elsewhere, and (2003b). Indeed, Bermúdez regards proprioceptive IEM as merely de facto.

from the fact that the concept of sensation makes no provision for identification of an owner.

In the case of memory, however, there is no comparable avowal: memory-judgements do not rest on sensation or any other introspectible state known with certainty. So a different explanation is required, one which appeals to the core conception of memory as retained personal information, as opposed to transmitted or new information conceived impersonally. Counter-examples against memory-based IEM assume an impersonal conception of information, detachable from the subject. This conception underwrites the possibility of q-memory – that is, personal memories allegedly originating in the experience of a subject distinct from the person who recollects them. A defence of the opposed personal conception of information thus forms the basis for arguments against q-memory.

In the case of proprioception there is a further, distinct explanation. The IEM of "My arm is raised" is explained by the qualified self-conscious knowledge criterion for bodily identity outlined below. It is also apparent from the defence of a conceptual holism between memory and personal identity, and between proprioception and bodily identity, which I now outline – that conceptual holism and IEM constitute a mutually supporting theoretical structure.

As noted in the Introduction, the treatments of memory and proprioception in this volume share a similar argumentative structure, yielding an integrated interpretation of self-consciousness as involving self-identification by a persisting subject. That common structure involves a response to the following dilemma in the case of memory and personal identity: either we define what it is for a subject to remember in a way that allows that the remembering subject could be distinct from the subject whose experiences or deeds are being remembered – but it then seems that memory cannot constitute a criterion for personal identity. Or else we define what it is for a subject to remember in a way that rules out that possibility – but then it seems that our definition must rest on a prior understanding of personal identity, and so the proposed account becomes circular. This chapter responds to the first horn of this dilemma by defending the claim that first-person memory-judgements exhibit IEM: that when I judge that I locked the door because I remember locking the door, my judgement cannot reflect knowledge that someone locked the door while at the same time involving a mistake about who it was that locked the door. The next chapter will respond to the second horn of the dilemma by showing that memory and personal identity form a conceptual holism, and that therefore there is no vicious circularity.

2.2 Memory-based immunity to error and the nature of personal memory

The phenomenon of IEM is exhibited by a wide range of self-ascriptions including those of sensation, perceptual experience, belief, intention, memory, and proprioception. The focus in the present volume is on two important cases which Wittgenstein omits from the *Blue Book* list, those of personal or autobiographical memory, and immediate awareness of bodily position and movement. Proprioception or bodily awareness yields immediate knowledge of bodily position and movement, and is a necessary basis for action. The IEM of bodily awareness is exhibited in such judgements as "My arm is raised"; when based on proprioception, there can be no question of continuing to maintain on the same basis that someone's arm is raised. In contrast, when anaesthetised, I might judge that my arm is raised when in fact it is someone else's that is.

IEM may be defined more precisely as follows:

> An assertion of "I am F", "I was F", "I o-ed", "I will o", and so forth, is IEM *if and only if*, were the subject later reasonably to doubt the assertion for whatever cause, it would be senseless for them to cite their original justification as support for the claim that nonetheless *someone* is F (or o-ed, or whatever.).

Hence, as Evans noted, IEM is exhibited by certain kinds of first-person statement, made on certain kinds of ground – memory rather than testimony for instance.[11] The definition requires development for such cases as Baldy and Mrs Gradgrind, since neither of them actually makes a judgement of the type "I fell out" or "I am in pain". However, their grounds for assuming that someone fell out, or that someone is in pain – whatever they may be – evidently leave open a possibility of misidentification that makes no sense.[12]

---

[11] Evans' account of IEM is found in his (1982), pp. 179–91, 215–25, 240–9. He describes the relativity of IEM to a "ground", as he puts it, in various ways. Evans claims that IEM is relative to a way of knowing about oneself (p. 218), but also that it is relative to "the subject's having ... or appearing to have information" about itself (p. 221; see also pp. 242 and 245), a formulation questioned below.

[12] Hence de Gaynesford's description of IEM as the "nonsense-question phenomenon" (2006, p. 53). His scepticism about the importance of IEM is addressed below.

It is important to realise that, in IEM cases, the retreat claim "Someone was F" makes no sense, *no matter what the cause of the doubt*. IEM must not be confused with the purely contingent immunity that obtains for a judgement based on testimony. I may be told that I had a serious accident when I was a child, but then discover that my informant is totally unreliable. I would then have no reason to believe that anyone had an accident, since the justification for the judgement is totally undermined by the informant's unreliability. In this scenario, one could not say that if the subject reasonably doubts the assertion for whatever cause, it would be senseless for them to cite their original justification as support for the claim that nonetheless *someone* was F; the implication follows only if they doubt the assertion because the informant turned out to be totally unreliable. If I doubt the assertion because my informant has sometimes confused me with others of similar appearance, then I may still believe that someone was F. In contrast, IEM arises only if the original justification cannot be cited as support for "Someone is F" when the assertion is doubted for any reason at all.[13]

By "original justification" I mean that which the subject would offer if asked to justify their assertion – their sincerity being assumed. The justification cites a continuous verb-criterion for personal memory-judgements:

> *Personal memory (1)*: A past-tense claim constitutes a personal memory-judgement only when the subject is prepared to offer the following justifications: "I remember o-ing … " or "I seem to remember o-ing … "; or "I remember (the) X o-ing" or "I seem to remember (the) X o-ing".

The former pair of justifications – "I remember o-ing … " or "I seem to remember o-ing … " – apply to my own actions or experiences. The latter – "I remember (the) X o-ing" or "I seem to remember (the) X o-ing" – apply to witnessings of the actions of others, or of events. For instance, someone asks how I know that I sold Samsonite suitcases, and I reply, perhaps sarcastically, "I remember selling them!", or more tentatively, "I

[13] IEM could be described as an internalist as opposed to an externalist or causal notion, since it concerns what the subject is justified in believing, as opposed to the causal basis for their belief in some actual state of affairs. I would argue further that the senselessness of the doubt constitutes IEM and is not, as Evans seems to think, merely a manifestation of it. The internalist-externalist distinction is criticised in Hamilton (2000).

seem to remember selling them". Or in the case of witnessings: someone asks how I know that my colleague was delighted that his monograph on self-consciousness had finally been accepted for publication, and I reply, "I remember him punching the air with delight when he received the letter of acceptance, and then – something he had never been known to do before – buying his colleagues a round of drinks".

To understand the ground for IEM-judgements, with its continuous-verb criterion, we need to understand the nature of personal memory. Personal or autobiographical memory concerns events which I have witnessed or experienced, where my judgements are *not* made simply as a result of receiving indications after the event – being told, or learning that it happened. I can be reminded or prompted, but to describe what happens in this way is to imply that my judgement is not made simply as a result of being told that something happened. Judgements that express purely factual memory of my past – also termed habitual memory of information, that which has been learned and not forgotten – in contrast are made simply as a result of being told, or reading, for instance, that I have done something. A tune that runs spontaneously through my head might count as a personal memory if it carried associations with an occasion on which it was heard, otherwise it would be factual memory. An ability to play the tune might be simply "remembering how". (Cases like this perhaps show that "factual memory" is better termed "habitual memory of information", but I continue with the briefer formulation.)

Although I may on occasion be unsure whether my knowledge of the past comprises personal or merely factual memory, the distinction remains a vital one. Indeed it is a datum of common-sense psychology – in Hacking's Wittgensteinian metaphor, "encoded in grammar" – and so it is presupposed by empirical enquiry concerning memory and not, as many psychologists tend to suggest, justified by it.[14] It is true that many personal-memory judgements take the form of simple past-tense claims, not preceded by the words "I remember…"; for instance, "I fell down the steps". But for a past-tense claim to count as a personal memory-judgement, the subject must normally be able to present it in the form "I remember falling down the steps…" or "I remember so-and-so falling down the steps…" – that is, using the *continuous verb-locution*. This locution implies the possibility of a spontaneous manifestation or willed rehearsal of the remembered events in the form

[14] Hacking (1995), p. 253.

of memory-images or memory-experience, accompanied often by feelings of pleasure, nostalgia, remorse, and so on.

There is another form of personal remembering which is more difficult to characterise: "I remember that tie you used to wear", "I remember Carlyle's house in Ecclefechan", "I remember the sinking feeling I had when I heard the news". The last of these can be paraphrased into a continuous verb-formulation, but the others do not refer to specific actions or experiences or events and so the paraphrase is not appropriate; nonetheless they are equally associated with rehearsal of imagery.[15] The most serious objection to the continuous-verb criterion is that languages such as Latin seem to lack it. What is essential is the possibility of a spontaneous manifestation or willed rehearsal of the remembered events in the form of memory-images or memory-experience.

The personal-factual distinction roughly corresponds to psychologists' distinction between "episodic" and "semantic" memory. Possibly as a result of Tulving's complaint about his discipline's neglect of "recollective experience" – earlier psychological work looked exclusively at distortions of *factual* memory, for instance of the contents of a document which the subject has read – psychologists now focus more on personal memory.[16] However, they seem to conflate it with other distinctions, notably between "remembering that", on the one hand, and "remembering how", or remembering to do something, on the other.[17] Personal remembering is neither "remembering how" nor "remembering that". The existence of an epistemic penumbra between personal and factual memory is stressed by some psychologists.[18] Now it is true that there is an uncontentious sense in which the exercise – as well as the concepts – of personal and merely factual memory of one's past is inter-dependent. Remembering times and dates often involves processes of inference and thus constitutes factual and not personal remembering, while past-tense claims are often based on an inference *from* memory; for instance: "What did you have for breakfast last Wednesday?" "Well, I was at home, and I can't remember when I last *didn't* have tea and toast for breakfast when

---

[15] The importance of such cases is emphasised by Byers (2005), who argues that the purpose of certain images is to evoke former times, that in remembering a certain image I remember an epoch. In such cases, she argues, there is no one-to-one relation between a time of remembering in which the image is "seen" and a previous moment or act.

[16] See Schacter (1989), pp. 692–3; Tulving (1983); Martin (2001).

[17] The latter is what psychologists often call "declarative" and "procedural" memory.

[18] See for instance Eacott (1998).

I was at home, so that's what I must have had." But the preceding examples do not undermine an essential distinction between personal and factual memory.

How then should one characterise the ground for personal memory-judgements, as cited in definition (1) above? Proponents of an *image theory*, which treats remembering as essentially a present experience from which one infers past events, often claim that putative memories are introspectively distinguishable.[19] For instance, John Pollock, acknowledging that many philosophers have denied a state of "seeming to remember", argues that "memory must provide us with beliefs about what we 'seem to remember' and then we infer the truth of what are ordinarily regarded as memory-beliefs from these apparent memories". He maintains that, by introspection, we can distinguish beliefs we have on the basis of memory from those we have on the basis of perception, "or for no reason at all". Or rather, we can distinguish *putative* memory-beliefs from other kinds: "I do not mean that we can tell introspectively whether we are correctly remembering what we take ourselves to be remembering".[20] This qualification avoids the implausible suggestion that memory yields infallible knowledge of the past. Image theorists must say rather that although I cannot be sure whether the event happened or not, I can at least be sure that if it happened, I know that it did on the basis of memory rather than because I received indications after the event that it happened.

The claim that we can distinguish putative memory-beliefs from other kinds by introspection, or any other kind of certain basis, seems mistaken. This is because I can ask myself "I o-ed, but do I genuinely remember o-ing or was I told that I had o-ed?" Goethe says in his autobiography that he is going to tell of things belonging to his childhood which he does not know whether he remembers or was told them. If he had been told, it would not be a case of personal remembering, but of merely factual memory. Consider the present writer's recollection of swallowing a ball bearing when a young child, and being taken to

---

[19] The theory is criticised in Hamilton (2003).

[20] Pollock (1987), pp. 51–2. It may be that Pollock is claiming that introspection reliably rather than infallibly tells us when a belief is based on personal memory. Ayer, I think, claims infallibility when he writes that when I decide that I really do remember something where I had previously been unsure, I pass "into a different phenomenological state: a state of which I am unable to give an informative description, but one that is familiar to anyone who has had the experience of having a memory 'flash' upon him" (Macdonald 1979, pp. 322–3).

hospital. As far as I know, this event actually happened. But it might have been something that I had been told; perhaps it seems to be a personal memory but in fact is not. Imagine that when I ask her about the incident, my mother denies that it happened to me, but has a vague recollection of it happening to someone. It would then make sense for me to say, "Perhaps this is something I was told. Maybe it didn't happen to me, but to a friend of mine". There are two interpretations of "I seem to remember swallowing a ball bearing": "but maybe I didn't", and "but maybe I was told".

So there are two kinds of apparent memory: "I'm not sure whether I o-ed" and "I o-ed, but I'm not sure whether I genuinely remember o-ing or was told that I had o-ed". With regard to very recent events, the latter kind of doubt generally makes no sense. Only in extraordinary circumstances – amnesia due to a stroke, for instance – could someone intelligibly say about a headache earlier the same day, "I don't know whether I remember having it, or was told that I had one". The senselessness of such uncertainty extends to salient though still fairly recent events in one's life.[21] "When I moved to Durham it was midwinter, but I'm not sure if I remember moving when it was midwinter or if I was told that I had moved when it was midwinter" would be pretty senseless if uttered by the present writer – unless, again, I suffer amnesia due to a stroke, and as part of my therapy am told such things about my past.

My suggestion is that "I seem to remember o-ing" is not a report of an introspectible state or "seeming", but rather a tentative claim about the past – tentative because of a specifiable doubt. There are no features of a putative memory-image by means of which the subject learns to distinguish it from the products of either testimony or imagination. Two questions now arise. How are memory-reports justified if not by introspective appeal to what I seem to remember? And what is the status of memory-images if they are not the basis for inference? In answer to the first question, "I was there, I witnessed it" makes explicit the basis on which the past-tense judgement was made. This is not a justification independent of memory, of course; it is itself a personal memory-claim that may be justified in many ways. Appeal can be made to the reports of other witnesses, and to material evidence: "There was a small earthquake here last night?" "How do you know?" "I was here, I felt it". "But how

[21] Appeal to the concept of "short-term memory" is unlikely to be helpful. It cannot be as short as the capacity required to follow a simple pattern of inference, however; this is a presupposition of any thinking at all, and not really memory at a personal level.

do you know it was an earthquake?" ... and so on. Where the subject's memory is a distinct one, these will be justifications for the audience only, not for the subject. In current terminology, such justifications are externalist in the sense of "not an introspectible or inner state", but internalist in the sense of "within the sphere of reasons".

The preceding considerations suggest a *direct knowledge account* of personal memory, according to which personal memory yields unmediated knowledge of the past. Such an account contrasts with the image theory, an "indirect knowledge account". On a direct knowledge account such as Thomas Reid's, the immediate objects of memory are past things and not present ideas: intentional objects, that is, objects of thought or judgement rather than awareness.[22] Just as fictional entities such as unicorns can be objects of thought, so can past objects. Thus remembering makes essential reference to judgement. To take Reid's example, to say that the smell of the tuberose is the "immediate object of my memory" is to say that my memory-belief is not the result of a process of inference, nor otherwise justified, from a present mental image, representation, or introspectible state. Personal remembering and the earlier perception share the same object. This is what it means to say that memory has past things for its objects.

A direct knowledge account exploits the analogy between memory and perception. That is, it regards personal memory-judgements as past-tense claims which are spontaneous and reliable, just as perceptual judgements are. If I am asked to describe the view from the window of the room I am in, I will mostly do so spontaneously, though partly via inference. (That square over there must be Red Lion Square, because I remember passing a street sign with that name on my way here, and so on.) In normal circumstances it would be absurd to question whether this really is what I am seeing; my judgements in this respect are authoritative. The reliability and spontaneity of perception is shared by memory. There is a human capacity spontaneously and reliably to report both present and past events. A spontaneous past-tense claim is one not made simply as a result of receiving indications after the event, whether statements or evidence, that the thing happened, and not based on an inference. The connection with perception is not just an analogy, however. Though with memory there is no possibility of re-focusing in response to questions – at least not in the sense of re-examining a transient phenomenon – personal remembering shares the same object as the earlier perception.

[22] Such an account was developed in Hamilton (2003).

Traditional direct knowledge accounts have neglected memory-images and memory-experience. My proposal thus constitutes a third way distinct both from such accounts and from the image theory. The image theory says that the memory-judgement is justified by memory-images – I am able to remember the event because I picture it. The traditional direct knowledge account says that the image is justified by the memory-judgement; that I am able to picture the scene because I remember it; and that the claim that the image is a correct representation of the event is itself a memory-claim. The third position advocated here is that the memory-image is an expression of the subject's knowledge of the events that they witnessed or experienced. (This position has affinities with the *no-priority account* of the Euthyphro dilemma offered in Chapter 3, below.)

Returning to the question of the basis for IEM, we can now see that definition (1) requires qualification. The basis cannot be described simply as apparent memory, since there are two kinds: "I'm not sure whether I o-ed" and "I o-ed, but I'm not sure whether I genuinely remember o-ing or was told that I had o-ed". Thus, to reiterate, "I seem to remember swallowing a ball bearing", can imply either "but maybe I didn't" or "but maybe I was told". In the latter situation it makes sense for the subject to say, "Well, it is at least true that someone swallowed a ball bearing". This is not personal remembering. In retreating to this position I come to believe that the impressions, if any, that I took to be memory-impressions of swallowing, distress, being taken to hospital, and so forth, are in fact products of imagination. Insofar as I entertain the retreat claim, I am wondering "Do I really remember this?" – that is, perhaps my apparent recollection is based on testimony. To reiterate, "I remember o-ing", or "I remember X o-ing", with the implication of memory-experience, is the hallmark of a justification involving personal memory. I can ask myself "Do I really remember something, or do I just believe it because I was told it?"[23]

One may suspect that if "I seem to remember o-ing ... " is the original justification, then the existence of apparent memory based on testimony undermines the claim of an IEM guarantee – since judgements

[23] In contrast – as discussed in Chapter 4 – no such uncertainty is possible in the case of present-tense judgements of bodily position. Judgements such as "My hair is blowing in the wind" have tokens grounded in sensation which exhibit IEM, and tokens grounded in visual perception which do not. But it makes no sense for me to say "My hair is blowing in the wind, but I'm not sure whether I feel this or see it".

based on testimony do not exhibit IEM. But it is important to realise that the subject's inclination to say "I seem to remember o-ing" is not a complete description of their epistemic situation. As noted earlier, such utterances are not reports of an introspectible state or "seeming". Rather, they are tentative claims about the past, where the tentativeness expresses a kind of doubt that must be specifiable. If I wonder whether the event really happened, the subsequent appearance of a doubt about whether I was told will constitute a *new* justification in the terms of the IEM definition. That is, I will not simply be saying "I at least seem to remember someone o-ing". (The point is elaborated in Section 3 below.) A full specification of the original justification is therefore as follows:

> *Personal memory (2)*: A past-tense claim constitutes a personal memory-judgment only when the subject is prepared to say "I remember o-ing…" or "I seem to remember o-ing…", or "I remember (the) X o-ing" or "I seem to remember (the) X o-ing", where any implied doubt is of the form "but maybe I (or the X) didn't".

The presence of the second kind of doubt does not undermine the earlier characterisation of "I seem to remember…" as a tentative assertion about the past rather than a report of an introspectible state. This kind of doubt involves a claim that is tentative about the source of evidence, and not directly about what happened, but I am unlikely to be confident that I know that something happened without knowing how I know it. To reiterate, "I seem to remember…" is always motivated by a doubt, and the kind of doubt must be specifiable.

The existence of apparent memory based on testimony has important consequences for the treatment of self-consciousness. The first of these is that there is no introspectible or avowable direct basis for memory-based IEM; "I seem to remember o-ing" is normally simply a tentative claim about the past. (We return to this issue later in the next section.) The second consequence is to confirm that the subject does not have infallible knowledge of its identity over time. There is no introspectible state, known with certainty, and underlying all sincere, spontaneous past-tense claims, since there is no introspectible difference between the two cases of apparent memory. Hence one should reject the powerful Cartesian intuition – "Cartesian" in the historically inaccurate textbook sense discussed in Chapter 4 – that the subject's stream of consciousness, extending back in time, yields infallible knowledge of identity.

This intuition exerts an unfortunate influence on both Locke's and Thomas Reid's conceptions of the self. Indeed in this connection Reid is more in agreement with his adversary Locke than may at first appear, since he writes that "the proper evidence [of a permanent self] is remembrance ... The evidence we have of our own identity, as far back as we remember, is totally of a different kind from the evidence we have of the identity of other persons, or of objects of sense. The first is grounded on memory, and gives absolute certainty".[24] We can now see that his claim is incautious. The subject has no infallible knowledge of its identity over time. The more restricted IEM guarantee captures the limited truth in Reid's claim of infallible knowledge. Paradoxically, the restricted immunity to error is often portrayed as Cartesian, and we can see from the present discussion why that portrayal is mistaken.

2.3 Justification of the immunity and its significance for self-consciousness

The centrality of IEM to self-consciousness arises from the understanding that having a first-person thought involves thinking about oneself in a distinctive way in which one cannot think of anyone or anything else. Intuitively, IEM is a hallmark of self-consciousness because the modes of knowledge that exhibit it – such as sensation, bodily awareness, personal memory, and intention – are those by means of which I am guaranteed to acquire knowledge solely of myself, if of anyone. In contrast, non-IEM cases, such as my knowledge that I weigh 12 stones, or – where the basis is visual perception – that my hair is blowing in the wind, are ways in which I can also know that someone else weighs 12 stones, or that someone else's hair is blowing in the wind. Hence the possibility of misidentification in such cases. Those who deny that IEM is a hallmark of self-conscious ways of knowing might argue that personal memory, like perception, yields knowledge of others; consider, for instance, the personal memory claim "I remember seeing him running out of the bank ... " Now it is true that memory involves knowledge, not simply of my own past states, but of events which I witnessed or experienced, including the actions of others. However, as is sometimes said, such knowledge of others is not "from the inside" – a misleading metaphor

[24] From this claim Reid derives one of the twelve first principles of contingent truth: that we know immediately "our own personal identity and continued existence, as far back as we remember anything distinctly" (Reid 1967, pp. 344, 345, 346, EIP III.iv; p. 445, EIP VI.v).

which means, simply, that the experiences or actions in question are not recalled as being mine.

First-person perceptual reports such as "I am looking at Marble Arch" are also IEM. But the presence of a diachronic dimension, and a synchronic bodily dimension when judging "My legs are crossed" on the basis of proprioception, makes memory- and proprioception-based IEM especially germane to self-consciousness, and it illustrates a neglected relation between "I" and bodily and personal identity. This bodily dimension of IEM is not much discussed by Evans and Parfit, while I believe the temporal dimension is misinterpreted by them in different ways. Evans claims that while memory-based self-identification spans past and present, Parfit treats it as essentially present tense.[25] However, this is not quite right. Where these writers differ is that for Parfit, memory-based self-identification seems to involve criteria, and can go wrong, whereas for Evans it is criterionless as regards the identity of the subject, and immune to error in that respect. So when Evans says that Parfit regards such self-identification as present-tense, when in fact it spans past and present, what he means is that Parfit fails to recognise that self-identification is *guaranteed* only in the present tense. In the case of proprioception the identification – or so-called identification – is synchronic but spans the bodily subject.

It is problematic to claim that the IEM modes of knowledge are those through which I am guaranteed to acquire knowledge solely of myself, however. Reference to "modes" or "ways of knowing" should not be taken to imply a *tracking model of self-identification*, in which the subject genuinely identifies a material object, itself; and "distinctively self-conscious self-ascriptions" or "capacities yielding IEM-judgements" might be preferable to "distinctively self-conscious ways of knowing". The issue is pursued in Chapter 6, which argues – against Evans's objections – that Wittgenstein was right to deny that criteria of recognition are brought to bear in IEM cases, and that there is genuine identification. IEM therefore qualifies the discrimination requirement of Evans and Cassam: that we cannot think about anything, even ourselves, without being able to discriminate it from the rest of the world, by suggesting a converse *non-discrimination requirement*: that I cannot think about anything without there being something – myself – which in central cases I do not need to discriminate at all. IEM also qualifies the apparent consequence of the discrimination

[25] Evans (1982), p. 246.

requirement that self-consciousness requires a substantive and accurate self-conception. It suggests a resolution of the apparent tension between a substantive account of self-consciousness, according to which the self-concept stresses the distinction between self-conscious and other ways of knowing about oneself, and a deflationary account focusing on the self-reference principle. If one can talk of avowals of intention to produce such a self-referring device – as in "I intend to produce a token of 'I'" – then as avowals they exhibit IEM; so the intention to self-refer self-consciously requires a prior IEM-exhibiting self-conscious capacity. Therefore IEM may be said to underlie the self-reference principle.

How the guarantee of IEM arises is a key question. The present account advocates a strong conception of IEM as an a priori truth for self-conscious creatures – a claim developed in the next chapter. But to reiterate, the explanation of its presence is not uniform across all varieties of IEM. In present-tense cases, the partial immunity to error of IEM-judgements is grounded in the complete immunity to error of avowals of sensation. The concept of sensation, in the range of cases under consideration, makes no provision for the identification of an owner, a feature carried over to present-tense IEM-judgements based directly on avowals of sensation. For instance, "My hair is blowing in the breeze", when it is IEM, is based on sensation; that ground, expressed as "It feels like my hair is blowing in the breeze", is completely immune to error. It is significant that present-tense IEM is generally regarded as uncontroversial, while the memory case, as we will shortly see, is not. No one regards Mrs Gradgrind's death-bed utterance quoted earlier – "I think there's a pain somewhere in the room, but I couldn't positively say that I have got it" – as intelligible, and there is a consensus that she and Baldy are trying to make identifications where it makes no sense to do so. Thus IEM is regarded as most secure when based on an avowable state. The cases of memory and anticipation do not seem to fit this pattern, however, while – as Chapter 4 will argue – proprioceptive judgements are not based on sensation. As we have seen, there is no avowable – that is, introspectible – basis of seeming to remember, since "I seem to remember…" is normally a tentative claim about the past. Moreover, the suggestion that IEM-exhibiting memory-judgements are based indirectly on states which were avowable at the time of the past event, if not at the time of recollection, still leaves open the possibility of misidentification. The explanation of memory-based IEM rests on the necessity of a personal conception of information, and the treatment of memory as

retained personal information.[26] The result is a non-empirical defence of IEM as a priori.

The defence of a conceptual holism between memory and personal identity, and between proprioception and bodily identity, in the next and subsequent chapters, constitutes a mutually supporting theoretical structure with IEM. To reiterate, the phenomenon captures the limited truth in the common assumption, held by Locke and Reid, that I have infallible knowledge of my own identity; it is the only substantive phenomenon underlying the unity of consciousness. It also constitutes the qualified truth in the memory criterion of personal identity, as the next chapter argues, while Chapter 5 shows how proprioception-based IEM reconfigures the debate over bodily identity. We will see that its presence in self-conscious bodily self-ascription means that IEM has anti-Cartesian implications, helping to undermine the traditional distinction between psychological and bodily criteria.

Objections to the phenomenon of IEM that arise from information garbling and from q-memory will now be considered.

2.4 Objections to the immunity: from misremembering

A common counter-example to the IEM of memory-judgements cites so-called misrememberings in which "information garbling" results in a misidentification of the subject.[27] A recurrent example is this: I am reminiscing with my colleagues about a philosophical meeting in which I made a brilliant objection to the speaker's thesis. However, none of us can remember any occasion on which I made that point. Then someone recalls Andrew Brennan making the very same point in a discussion which bears striking similarities to the one I have described. I am persuaded that this indeed was the discussion in question, and that I have misidentified the person who made the objection. "I made a brilliant objection" thus seems to be a misremembering – a personal

[26] Brinck argues that IEM is based on judgements that are not reports of states of oneself, but a direct expression of these states: "The non-conceptual content that results in IEM cannot come from another source than myself, and it cannot be used as a ground for ascribing a state to a receiver distinct from the source. The experiencing subject and the judging subject are one and the same" (Brinck 1997, p. 136, also p. 132). But she does not discuss the details of the memory-case; moreover, I would prefer to speak of avowable states that give rise to avowals, rather than non-conceptual content.

[27] A common counter-example in my experience of presenting papers on the topic.

memory-judgement with an incorrect detail, mistaken due to an error in identification. Many commentators would say that my reminiscence was based on information that was garbled, resulting in a misidentification of the maker of the objection as myself. Thus memory-judgements are not guaranteed IEM.

Crucially, the details of the objection are left vague. When they are brought into sharper focus the objection loses its force. For the objection to be sustained, the subject's initial claim has to be fairly tentative; a confident claim of a distinct memory would be completely undermined by the subsequent testimony. Nor should it simply be a case of jogging one's memory; I do not say, "Yes, of course, I remember now, it was Brennan". Rather, I still do not remember that it was Brennan, but take it on testimony that it was him. Hence all I am meant to say is "I at least remember that someone made the objection". However, this last formulation requires a further refinement, which undermines the case against IEM. I have presented the original claim as "I made a brilliant objection" and the retreat claim as "I remember that someone made the objection". But in fact the continuous-verb locution – whose importance was emphasised earlier – is required here: "I at least remember *someone making* the objection".

A dilemma now arises for the critics of IEM. "I remember that someone made the objection", or "I remember that those words were spoken", are not sufficient for personal memory. In order for the example to constitute an objection to IEM, the speaker must be represented as saying "I at least remember someone making the objection" – that is, they must be credited with a recollection of the experience of the event, not merely factual memory that it occurred. But this is not a credible claim for them to make, since for it to be true, the subject's memory-experience would have changed to accommodate new evidence – thus invalidating its status as memory-experience. This point applies even when the initial claim was just a vague recollection of making such an objection.

But surely, it may be argued, someone could be more sure of the objection's having been made than of who the speaker was? They may mention the choice of words, the stage of the proceedings, the tone of voice, degree of emphasis, the speaker's reply, and so on. This amount of detail, the argument goes, justifies the use of the continuous-verb form "I at least remember someone making the objection" – and not the merely factual formulation, "I remember that someone made the objection". To this point, the reply is: the continuous-verb claim might be justified where initially I had thought that Brennan made the objection, and now I am not sure who did. But that claim would not be

justified if I initially thought that it was myself who had made the objection, because the detailed descriptions are in part third personal, and I could not adopt the appropriate third-personal stance towards actions which are my own. How, for instance, could I report on my own tone of voice in this way? The significance of this point will become clearer shortly. My conclusion is that, in the terms of the definition of IEM presented at the start, the evidence from testimony constitutes a new justification and so leaves the IEM status of the original judgement unaffected. Since I have now been told that someone made it, the original justification has been superseded by a justification that is inferential, non-spontaneous, and hence non-memory-like. Without this additional testimony the idea that, if it was not myself, then it must have been someone else who made the objection, would just be a stab in the dark.

It might be said that, given the general reliability of memory, I have reason to believe that something like the making of the objection did happen. But I will have no clue as to which detail is mistaken. I might as well say, "I at least remember that I said *something* at that meeting". Perhaps I do have a specific reason to wonder whether it is the alleged detail of identification – of who it was that made the objection – that is mistaken. There are cases of motivated misremembering, in which the subject's role becomes exaggerated, and maybe this is one of them. If I know that I am prone to such self-inflating, or self-protecting, gestures, I may perhaps wonder whether the objection was made by someone else. But these responses continue to miss the point. In each case, my inclination to make the past-tense claim is treated as a basis for inference, when taken in conjunction either with the general reliability of memory, or with knowledge of my own tendency towards misappropriation. Yet personal memory has a non-inferential character and so the result in these cases will not be a personal memory-like judgement. Again there is no license for the retreat claim "I at least remember someone making the objection".

The defence of IEM in the face of the "brilliant objection" objection is quite consistent with holding that the making of the initial claim – namely, "I remember making a brilliant objection" – is not completely accidental. Even if it does not constitute a misremembering – a personal memory that is incorrect in detail – it might not be a complete coincidence that the subject comes out with an apparent reminiscence which tallies in important respects with those of other witnesses. To say that the apparent reminiscence, while not a misremembering, is no accident, is to suggest a distinctive category of defective memories. In order to

explain this, we need to examine the different possible errors to which memory-judgements are subject.

These are:

i) misremembering or false-in-detail memory
ii) completely false memory
iii) judgements based on testimony

And also, the new category:

iv) not a personal memory, and not based on testimony, but not completely false. (In the case of the brilliant objection, the retreat claim is based on testimony, but the original claim is assumed not to be.)

Completely false implies that nothing like the events reported ever happened – "completely false memory" is short for "sincere, attentive, but completely false memory-claim". Such memories are necessarily rare. Where memories are mistaken, they are almost invariably *false-in-detail*: in ordinary language, misrememberings. The neglected contrast between false-in-detail and completely false memories is illustrated by Ian Hacking's discussion of False Memory Syndrome, where he contrasts an alleged misremembering – a patient apparently recalls being abused by her uncle when in fact it was her father who was the abuser – with a "completely false" memory of the same event, where no abuse of any kind happened.[28] The contrast between false-in-detail and completely false memories is imprecise and context-dependent, but as in the case of perception, it is a vital one. "Imprecise and context-dependent" does not mean that the distinction is merely a matter of degree; it is a categorical difference, though opinions about it may reasonably differ in context.

The claim that mistaken memory-claims are normally merely false-in-detail is not based on psychological research, but rather is how claims such as "Your memory is mistaken" or "Your memory deceives you" are understood. If the subject is assumed to be sincere, we are inclined to think that their memory-report is at worst false-in-detail; this is just what the reliability of memory amounts to. It can happen that what I think I remember turns out to be something I dreamed vividly, or

[28] Hacking (1995), pp. 258–9; he refers to the two cases as "contrary" and "merely false". I would argue that misidentification of the perpetrator is too significant to count as a misremembering, but the distinction still stands.

saw in a film, or was told. But such completely false memories are rare, and special circumstances are needed to explain them.

The general reliability of memory must be distinguished from the particular guarantee of IEM – as J. S. Mill fails to do in his discussion of the self. When he explains why he now believes his series account of the self to be inadequate, he claims that "a remembrance of sensation ... involves the suggestion and belief that a sensation, of which it is a copy or representation, actually existed in the past". That is, memory-beliefs are reliable – or, as Mill puts it, memory is a form of intuitive knowledge. However, he goes on to say that the phenomenon of memory "[cannot be] adequately expressed, without saying that the belief [it includes] is, that *I myself* formerly had ... the sensations remembered".[29] But this is a different point: that the past sensation, assumed to have existed, belongs to myself.[30]

To return to the brilliant objection example. To say that the initial reminiscence was no accident is to assert the counter-factual "If I had not witnessed the objection being made, I would not have come out later with the apparent reminiscence". Those who hold that this counter-factual is true often assume a seductive picture involving a prior, determinate information-link between the discussion and the subsequent mistaken reminiscence, in the course of which the information has somehow become garbled. The assumption, often made unwittingly, is a decisive philosophical move. I will argue that one should not assume that the participants' investigation is tracking, or failing to track, any such information-link, since the concept of information it appeals to is impersonal or neutral. This impersonal conception should be rejected. "There is an information-link" means just that "We infer that the subject retained some knowledge or information concerning the earlier event, though possibly in a confused form". The testimony of others, and collateral evidence, is required in order to decide whether knowledge or information is retained.

Describing the reminiscences of the brilliant objection as "information-garbling" therefore involves a questionable metaphor. Unlike the alleged reminiscence and collateral evidence and testimony, the existence of a prior, determinate information-link is not a datum, but the product of an inference. Sometimes the inference to

---

[29] *An Examination of Sir William Hamilton's Philosophy*, in Mill (1979), p. 194, my italics.

[30] The reliability of memory is discussed at greater length in Hamilton (2003).

an information-link seems compelling, as in this example from Hoerl. A Korsakoff's patient rapidly forgets being pricked with a pin by the experimenter, but when the experimenter reaches out for her hand later, she withholds it, saying, "Doesn't one have a right to withdraw her hand?" It is plausible to postulate an information-link here: the subject would not have said what she did if she had not been pricked with a pin earlier. However, as Hoerl rightly comments – in opposition to the views of some psychologists working on amnesia – past events may influence us without it being said that we remember them.[31] Information retention is not sufficient for personal remembering. To say that the apparent reminiscence is no accident suggests a connection at the information-processing level, where "information" does have a technical and impersonal or sub-personal sense. But even so, it does not follow that there is a determinate information-link of the kind suggested; to reiterate, the informational connection is not sufficient for personal memory, which is why "I at least remember someone making the objection" makes no sense.

I am suggesting that the senselessness of "I at least remember someone making the objection" should be explained by appeal to the IEM of personal memory-judgements. However, there is a further, and fundamental, objection. It may be argued that even if the "brilliant objection" objection does not show that the IEM guarantee is illusory, it may still be just an instance of a larger and less interesting phenomenon. The resulting *deflationary account of IEM* treats it as just one kind of error which is too large for a misremembering. On this view, it is the general feature "too large an error for a misremembering", and not any particular immunity concerning self-identification, which explains the senselessness of the retreat claim "I at least remember someone making the objection". The conclusion is that IEM is merely an illustration of a more general feature of personal memory: that, as in the case of perception, when the subject seems to be mistaken in more than a matter of detail, doubt is cast on their claim in its entirety. If, for whatever reason, I come to doubt some significantly large feature of my memory-claim, it will make no sense for me to continue to maintain that I personally remember anything at all of the events concerned. Mistaking myself for the objector is too big a mistake for a misremembering; claiming

[31] Hoerl (1999), p. 234. Another example is the way that brain-damaged, profoundly amnesiac patients can continue to develop skills such as operating a personal computer, despite avowing at the start of each training session that they have never used one before.

that Millar made the point when it turns out to be Brennan most likely would not be.

The IEM case would then be assimilated with examples such as the following. I seem to recall seeing British comedian Tommy Cooper at the London Palladium, wearing a Panama hat, but my friend assures me that the only hat he wore was his trademark fez, and on reflection I believe his testimony. It would be odd for me now to maintain, "Well, I at least remember him wearing some kind of hat". Possibly, if I overheard someone saying "Tommy Cooper's hat is ridiculous", I might say, "I remember that he wore a hat – someone made a comment about it". But again the error seems too significant for a mere detail. Contrast an uncontroversial case of misremembering – where I initially claim that he was wearing a fez but then come to believe that it was a different but similar-looking Middle Eastern hat.

IEM should not, I think, be assimilated to such cases. The identity of the speaker is indeed "too big for a misremembering", but it does not follow that IEM is simply an instance of the more general phenomenon that some errors are too big for a misremembering. IEM of self-ascription captures the Distinctness Principle, the idea that one thinks about oneself in a distinctive way in which one cannot think about anything else, which as argued earlier is essential to self-consciousness. Memory-based IEM is one of several varieties of self-ascriptive IEM, such as sensation, proprioception, self-location, and propositional attitudes. "Who it happened to" – when the event possibly happened to myself – has a different status to other elements of the memory-claim, and wider ramifications. The reason for this may be that although, as argued earlier, the IEM of personal memory has no direct basis in avowals, the "brilliant objection" example shows that there are aspects of recollection that could not be reported in a third-personal way: in this case, for instance, the speaker's tone of voice. Such aspects are those capable of being expressed in an avowal at the time in question. The continuous-verb form definitive of personal memory thus captures in a unique way the first-person viewpoint of avowals.

2.5 Objections to the immunity: from q-memory

Unlike the objection from information garbling, the objection to IEM from q-memory usually takes a science-fictional form. Advocates of q-memory generally assume one or both of the prevalent image and trace theories of memory, though as we will see q-memory does not strictly presuppose these. (The theories can, I think, be held simultaneously.) The *trace theory* is often assumed without question, as when Parfit describes one person's

memories being "copied" into the brain of another.[32] This theory locates the essence of memory in brain-traces, and makes q-memory seem more plausible because it characterises the causal connection between the witnessing and the remembering in neurophysiological terms. The identification of the witnessing subject with the recollecting subject is eroded, making it seem possible that such a connection could obtain between a witnessing had by one subject and a recollection had by another. The present discussion aims to refute such speculation.

In addition to resting on the trace theory, q-memory also assumes the image theory, which – to reiterate – treats remembering as essentially a present experience from which one infers about past events. Cassam for instance supposes that q-memories "come to us in the first-person mode of presentation; they present themselves as concerning the rememberer's *own* past experiences. However, the *belief* that they are about [my] own experiences is, in principle, 'separable'".[33] The image theory separates experience and belief or judgement, treating memory as essentially a content or image distinct from what the subject is inclined to judge. A more plausible characterisation would be: I am inclined to say that I remember o-ing, or remember someone o-ing, but later come to believe that I am mistaken, while judging that my apparent experiences bear a resemblance to what I imagine the past experiences of someone else to be.

Q-memory yields the most common line of objection to IEM. In the literature, *q-memories* are defined as phenomenologically identical to and caused in an appropriately similar way to ordinary memories, but originating in the experience of a person other than the remembering subject.[34] While the brilliant objection scenario is premissed on information-garbling, q-memory is based on *information-transfer*. *Mere q-memories* are those which concern events that happened to someone else; nothing like the event in question happened to the subject who q-remembers it. The brilliant objection, in contrast, hypothesises that this

---

[32] "Jane has agreed to have copied in her brain some of Paul's memory-traces. After she recovers consciousness in the post-surgery room, she has a new set of vivid apparent memories" (Parfit 1984, p. 220). There is a plethora of comparably science-fictional examples in Parfit's writings.

[33] Cassam (1997a), p. 41.

[34] Parfit (1984), Chapter 11, especially pp. 219–22; Shoemaker (1984b), especially pp. 41–3. The term "quasi-memory" may have originated with Samuel Alexander, to denote an image "contemplated as somewhere in Time, but the position of it as a whole in Time is not dated" (1920, pp. 119–20). This is evidently not Parfit's notion, and Alexander seems to advocate a kind of direct knowledge account.

was an event that concerns my past, in that I witnessed it, even if I did not perform the action concerned, and that later I misremember it. If this kind of error were systematic – if, for instance, it was always the actions of Andrew Brennan which, after witnessing them, I misappropriated – then the phenomenon might be close to what is termed q-memory.

Q-memory is associated with *reductionism about persons*, which is advocated by Parfit, and denies that there is any genuine IEM guarantee. Reductionism in this sense holds that there is a level of thought about oneself more fundamental than that involving use of the first person; that is, the world can be described fundamentally in impersonal terms. On this view, a "complete description of the universe" refers to brains, bodies, and experiences, to which persons are reducible: "A person's existence just consists in the existence of a body, and the occurrence of a series of thoughts, experiences, and other mental and physical events".[35] A critique of reductionism concerning persons should begin by focusing on the pragmatically self-defeating nature of its statement, and on the implausible supposition of a "complete description of the universe". This volume is not concerned with a sustained critique of reductionism, but rather focuses on its erroneous espousal of q-memory and rejection of IEM. Reductionists hold that the remembering subject only *assumes* that they are identical with the person who had the experience, an assumption that is justified only by the contingency that q-memory is not prevalent.[36] I am concerned to reject this position, and present an anti-reductionist defence of IEM, thus elucidating the idea of the subject as a persisting entity.

Commentators distinguish strong and weak concepts of q-memory. *Strong q-memory* is, as Evans puts it, a faculty whose "informational states … announce themselves … as *merely* q-memories, so that it seems to the subject that someone or other o-ed without its in any way seeming to him that *he* o-ed". Such a faculty, he writes, is possible only on the strongly Parfitian assumption that it is not part of what I seem to remember about an experience that I, the person who now seems to remember it, am the person who had it.[37] It is hard to know what to make of this assumption. We often make claims of the form "I know that

[35] Parfit (1995), p. 16.

[36] See, for instance, Parfit (1971), p. 15.

[37] Evans (1982), p. 248; the assumption is made by Parfit for instance in his (1971), p. 15. This is the kind of q-memory advocated by Noonan (1989), pp. 182–5. Q-memory is usually taken to be a peculiar kind of personal memory, but one could equally postulate that the subject q –factually remembers what someone else had learned or was told.

someone or other o-ed but I can't remember who", or even "I remember someone o-ing but I can't remember who" (though not "I distinctly remember…"). But a strong q-memory is meant to include myself among those who possibly o-ed, hence "I seem to remember someone arriving in Bournemouth on holiday, though I am unsure whether it was myself or someone else".

We do not need to pursue strong q-memory, however, because it is subject to the same critique as the apparently more plausible *weak q-memory*. Evans defines it as follows:

A subject (weakly) q-remembers an event if and only if:

i) he or she has an apparent memory of such an event.
ii) that apparent memory embodies information deriving from the perception of that event by someone who is not necessarily the person who apparently remembers.[38]

Evans's definition is satisfied by apparent memory based on testimony, and so a further condition is required to bring out that q-memory is meant to be a peculiar kind of memory:

iii) the subject was not told or did not otherwise receive the information in a non-memory-like or extraneous way.

An example of weak q-memory would be my seeming to remember going on holiday to Bournemouth. In contrast to strong q-memory, it is not left unspecified who went on holiday. Although it turns out that I never had a holiday there, the apparent recollection is not completely false, but – so proponents of q-memory claim – embodies information from someone else's holiday in Bournemouth. So I merely q-remember.[39]

One should note that in his definition of strong q-memory, Evans, like Parfit, is in thrall to the image theory. Both writers assume a genuine difference in content between strong and weak q-memories; indeed Evans believes that these varieties of q-memory are distinguished by their respective compatibility and incompatibility with IEM. Each writer evidently sees a distinction between the way in which the q-memories "announce" themselves – their introspectible content – and the subject's

[38] Evans (1982), pp. 247–8.
[39] Henceforth "q–memory" = "mere q–memory", that is, it is assumed that the q–memory comes from a different subject.

belief about the past event. This distinction is denied by the direct knowledge account of memory, which holds that the only real difference between the two kinds of q-memory lies in whether or not the subject regards the phenomenon as a completely false memory in the sense above, or else is convinced that they are experiencing a peculiar memory-phenomenon. Once the subject comes to believe that these weak q-memories originate in the past of someone else, they will indeed "announce themselves" as merely q-memories.

While conceding the possibility of weak q-memory, Evans denies that it could constitute knowledge of, as opposed to information derived from, the past of someone who was not myself; if it served as the basis of knowledge of that person's past, he believes, it could not count as a normal operation of memory. But the resulting position still grants too much to proponents of q-memory. Both Evans and McDowell seem to agree with Parfit that, from a scientific psychological vantage point, there could be (weak) q-memories – that is, information originating in the past of someone other than the rememberer. I wish to reject any such fundamental vantage point, and thus undercut the Evans–Parfit dispute.[40] In doing so, two issues must be addressed. This chapter focusses on the impersonal concept of information, and the use to which it is put in information models of memory – a use of which Evans and McDowell are insufficiently critical. The next chapter discusses how Evans and his sympathisers favour *materialism concerning self-consciousness*.

2.6 Rejecting q-memory: concepts of information

Even those sympathetic to q-memory would accept conceptual limits to its prevalence, since according to any but the most strictest bodily continuity criterion, the existence of q-memory-links has implications for personal identity. Hence *distinctness of subjects is not generally a datum in q-memory scenarios*. In a scenario of pervasive q-memory-links, distinctness of subjects becomes an interpretation and not a datum; one has to show that the subject which had the experience and the subject which allegedly later q-remembered it really were distinct. A dilemma then arises: is the correct description, "X has q-memories originating in experiences had by Y", or, for example, "X and Y are the same subject (one

[40] See Parfit (1984), pp. 221 and 516; Evans (1982), pp. 244–5; McDowell (1997). Parfit is responding to Evans's objections.

Lived Body, two mere or physical bodies)"?[41] If the latter description is adopted, there would be no q-memory-link.

These considerations do not furnish a thoroughgoing critique of q-memory, since the intuitive distinctness of subjects in some such scenarios *is* a datum: limited occurrence of q-memories linking distinct subjects is still possible. We must therefore question the materialist or functionalist assumptions which commonly underlie this latter possibility. (Later we consider the possibility of limited q-memory without such assumptions.) Proponents of q-memory mostly assume a causal or functionalist definition of personal memory, regarding it as an internal informational state occupying an appropriate causal role, connecting past experience or witnessing and present recollection. The assumption of a determinate information-link, criticised earlier in connection with information-garbling, underlies the common science-fictional scenarios of memory-transplant and person-replication. Without materialist or functionalist assumptions, these scenarios would not be possible. However, both the assumptions and the science fiction seem to be accepted by opponents of q-memory such as Evans and Cassam, suggesting that a deeper critique of q-memory than theirs is possible.[42] To reiterate, there is a common assumption that one can postulate an information-link between distinct subjects without first establishing a way of knowing.[43] This assumption underlies science-fictional accounts both of memory-transplant, and of the more ambiguous concept of memory-transfer, which does not have to involve the physical relocation of brain tissue.

To develop the point made in discussion of the brilliant objection: a determinate information-link is not a datum. The data are the alleged reminiscence, plus any correlations that may be drawn through testimony and material evidence between the content of what are simply – by ordinary standards – erroneous past-tense reports, and the past of an individual distinct from the maker of the report. "There is an information-link" means simply "People infer from such phenomena to facts about the past of another, and reliably so". Until the existence

---

[41] The contrast between Lived Body and mere body is pursued in Chapters 4 and 5. The claim of indeterminacy is denied for instance by Noonan, who confidently predicts that in "other possible [q-memory prevalent] worlds … our concept of a person still has unproblematic application" (1989, p. 191).

[42] Evans (1982), pp. 241, 244; Cassam (1997a), p. 41.

[43] The tendency is found in Evans, and also in Martin (1995). A common alternative formulation to "way of knowing" is "belief-forming process", but in both cases there are serious problems of individuating the way or process.

of such correlations leads to the postulation of a way of knowing, there is no as yet unrecognised information-link. Proponents of science-fiction thought-experiments are therefore wrong to assume that it is fixed in advance whether certain phenomena indicate the presence of a way of knowing. Moreover, even if a way of knowing were postulated, further conditions would need to be satisfied for it to count as a memory-like one.

These criticisms of determinate information-links assume that information is a personal notion, an assumption that, insofar as the issue is recognised at all, is likely to be contested. We must therefore defend the *personal conception of information*, which claims that its presence implies a recognised way for a subject to gain knowledge – direct, or theoretical and inferential – by means of it. On this conception, information does not just physically lie around in the universe waiting to be acquired; it is only information *for someone*.

There are different variants of the personal conception. A *strong personal* conception says that the ordinary sense of information must involve a sender who intends to inform as well as a receiver who is (potentially) able to interpret – that is, the conception takes the cognate "inform" seriously. This conception makes a sharp distinction between information and mere evidence. Thus when Robinson Crusoe infers from the footprint in the sand the presence of another human, he is relying on evidence of the existence of another islander, but not on information about him, unless Man Friday had come to know of or suspect Crusoe's presence, and intended to let him know of his own by leaving conspicuous footprints (or, bizarrely as it turns out in Defoe's tale, just one footprint). This conception is too strong; the footprint clearly does yield information about the size of Man Friday's foot.

More plausible is a *weak personal* conception which says that information should be interpretable; an interpreter of information is required, but a sender is not. This weaker conception equates information with being informed, rather than with communication. Fossils, on this view, do not constitute information about extinct species until they are recognised as containing their remains; tree rings do not constitute information about the age of trees until we know how to calculate it by using them. The theory of tree growth shows how slower growth in winter leads to the formation of denser bands in the tree trunk. But such information could not be said to exist prior to the theory's development; to describe it as "potential information"

implies an unacceptable explosion of information attributions.[44] It might be argued that the aftermath of "the perfect crime" contains no information about the identity of the causally responsible agent. According to the personal conception of information, however, there is no determinate answer to the question "What information does this contain?", and so no guarantee that the crime was perfect; for instance, advances in DNA testing have led to the solution of what might previously have been regarded as perfect crimes.

Information has been discussed in contemporary philosophy, but rarely analysed in depth. Dretske's *Knowledge and the Flow of Information* is a partial exception to this neglect, and endorses a personal conception of information for what he calls its "nuclear" sense: "Information is what is capable of yielding knowledge".[45] However, Dretske distinguishes this concept from one involving meaning, from which a technical sense of information-content, based on the mathematical theory of communication, may be developed. According to this sense, information exists independently of the minds of interpreters or of an intentional stance; hence Dretske argues for a naturalistic account of the development of conceptual or semantic content from informational content.

This account is problematic because a personal conception of information neither affirms nor denies that without the possibility of agents acquiring knowledge of them, events are "intrinsically meaningless". Indeed it is not clear what "intrinsically meaningless" amounts to, except perhaps the denial of a pantheistic or animistic world view. However, the impersonal conception clearly affirms that information physically lies around in the universe waiting to be acquired, and so does not have to be information *for someone*. Evans's concept seems intermediate between Dretske's and the personal conception. He defends the "belief-independence" of information and argues that "we cannot speak of an information-link when there is any process of *inference* on the part of the subject"; yet, as discussed earlier, he also seems to allow a

[44] While the strong personal conception contrasts information with evidence, the weak personal conception might be said to equate them, in that interpreters, though not senders, are required for both, though since evidence is always evidence for something, it might be regarded as essentially propositional, while information is not.

[45] Dretske (1981), p. 45. Dretske's account is usefully discussed by Adams (1991). For further discussion see also Deacon (2012), Chapter 12.

q-memory information-link, while arguing that it could yield knowledge only inferentially.[46]

The *impersonal conception of information* is a quantitative concept. It denies that, for something to count as information, subjects have to acquire knowledge by means of it. The conception originates with engineer C. E. Shannon's mathematical theory of information, originally applied to the transmission of voltages in telecommunications because it is through this means that speech may be communicated.[47] The resulting information theory postulated quantitative concepts of information-content, redundancy, and entropy. "Information" has thus acquired two meanings: one technical, the other non-technical. (Other examples include "work" and "energy", though these do not seem to cause such fundamental philosophical problems.) The impersonal conception says that the core sense of "information" is impersonal, and naturalistic philosophers of mind have favoured it.[48]

How might information-links be characterised on an impersonal conception? Among the latter's proponents, a purely causal characterisation connection is generally discredited. Dretske, for instance, argues that "Just as there may be little, if any, information contained in an effect about its cause, an event may carry information about events to which it stands in no regular causal relationship". He develops a notion of "information-bearing" in terms of having the function of nomically depending on something, which is a much more agile relation than a purely causal relation.[49] "Information-bearing" is also defined in terms of covariance, as follows: if A covaries with B, and the covariance is non-coincidental, then A bears information pertaining to B.[50] The problem with equating the bearing of information with covariance is

[46] Evans (1982), p. 146n. The distinction between information as existing "in the world" and as acquired by an agent is also discussed in Devlin (1991), p. 16.

[47] Shannon's original paper appeared in the *Bell System Technical Journal* in 1948; details of his theory are given in Dretske (1981), p. 237. The issue of tree rings as information is discussed in Lowe (1996), pp. 119–20.

[48] As have popularisers of science such as Jim Alkalili, who argues that "information is not just about human communication, it is woven very profoundly into the fabric of reality" (BBC 4, *Order and Disorder*, episode 2, 23 October 2012).

[49] Dretske (1981), quotation from p. 31.

[50] I have even heard it claimed that the "informational content" produced by excessive alcohol consumption results in vomiting! The idea that water running downhill is doing "information-processing", presumably by working out the line of least resistance, is another example, used to humorous effect in Searle's (1984), p. 50.

that it yields no criterion for distinguishing information from misinformation, and – to use the terminology of its proponents – no way of determining which chain in a causal process the information is to count as representing.

This is a key issue. The impersonal conception cannot accommodate the idea of misinformation, since the latter involves the intention to deceive; it is deliberately false, whereas false information may be accidentally false. Random static on a phone line may cause my fax machine to produce the message "I'll be at the station at 7. Dave". I may as a result believe that my friend will arrive at seven, but it would be wrong to say that I had been misinformed. Nor could an impersonal definition sustain a distinction between true and false, or reliable and unreliable information. All of this is conceded by Dretske, whose response is that the ordinary concept does not encompass misinformation either.[51] He argues that since information is what is capable of yielding knowledge, and knowledge requires truth, so too does information; so although we do speak of false information or misinformation, these are not part of the core concept. Now although Dretske is right to suggest that the personal concept of information is defined in terms of the capacity to yield knowledge, it is wrong to assume that false information is not information. "Inform" carries the implication of "correctly inform". But someone can misinform me that the meeting is at 2 pm, when it is in fact at 10 am, because they do not want me to be there. It would certainly be true to say that I received from them the information, which turned out to be misinformation, that the meeting was at 2 pm.

We need to ask whether an amended definition of information is necessary in order to accommodate misinformation, and indeed whether it can do so. The personal concept could be amended to say that information involves a recognised way for the subject to gain knowledge, or justified (evidence-based) but false belief, by means of it. The addition of the "justified (evidence-based) but false belief" clause might seem advisable for another reason: one is not guaranteed to acquire knowledge from genuine information, since there is always the possibility of misinterpretation. These are not reasons for amending the original definition, however. It is at least possible to gain knowledge from genuine information, whereas any true belief gained accidentally from false information would not count as knowledge. The amendment does not seem sensible in any case. It is better to say: information, when true, is capable of

[51] Dretske (1981), p. 45.

yielding knowledge for a subject. Thus problems over misinformation are resolved, and Dretske's criticisms of the personal conception are mistaken.

It might still be urged that the impersonal conception is perfectly coherent, and justified by the following mechanical example: when a camera shutter is opened, information is transmitted from the scene to the film.[52] However, this description is justified only because camera and photographic image are artefacts yielding information in a personal sense. Their primary use, simply put, is to record events in fixed images for people to look at afterwards. Contrast a natural photographic image produced by chemicals that have seeped into a pool of water at which elephants drink. Assume that the ponderous elephants remain motionless at the pool long enough for an image of them to become fixed at its bottom. Does the image contain information concerning elephants drinking? For humans, perhaps, and maybe for other creatures who may be startled by it, thinking that there are elephants in the pool – though this looks more like a case of misinformation.

In the technical sense, and for a particular range of cases, there may be a quantitative answer to the question "What information does this object contain?" But in the personal sense there is no unique answer for all observers. The information which a photograph contains is open-ended, and relative to an interest and existing state of knowledge; it will yield different kinds of information to a holiday snapper and a photographic expert. To return to an earlier example, it cannot be specified in advance of new methods of DNA testing how much information about a criminal's identity the crime scene contains. Normally the claim that information is present is accompanied by some understanding of what it concerns, though there is a use for the claim "There's a lot of information here, but I've got no idea what it's about" by the discoverers of the Rosetta Stone, for instance, trying to decode its ancient inscriptions. These scholars of course assumed that the inscriptions constituted communication by an agent or agents – that they are intentional objects. So the centrality of the personal sense of information is not undermined.

I have denied that the impersonal concept of information is basic, and questioned its viability or coherence. The impersonal conception is, I believe, at best a derivation from, or kind of metaphor for, the primary

[52] The example is found in D. M. MacKay's entry on "Information Theory" in Gregory (ed.) (1987); he then distinguishes "structural" from "metrical" information-content.

personal conception. One may still wish to stipulate a technical definition, and in genetics, artificial intelligence, and in information-processing models in human perception and cognition, this may be justified, for instance when the genotype is described as containing genetic information or a genetic code. (The concept of encoding raises comparable problems to that of information.) If the genotype is a plan or blueprint, it is one which has no agent to put it into effect. There is information in the ordinary, personal sense only if the geneticist, on examining the genotype, can predict the nature of the phenotype, and make claims such as "And this tells us he will have blond hair". Teleological or purposive explanations in biology suggest agency, which their proponents know is literally false; the code "tells" the cells their function only in a metaphorical sense.[53] The problem is that although it is possible to define a technical concept and call it "information", the latter concept is conflated with the ordinary one, whose primacy becomes obscured.

The preceding defence of the personal concept of information has shown that the concept has no logical priority over that of a knowing subject of experience. One cannot postulate information-links and then decide whether or not it is persons that they link. This is a fundamental line of objection to the concept of q-memory as ordinarily presented. However, there are other ways of presenting it which may be more persuasive, and should now be considered.

2.7 Rejecting a more conservative presentation of q-memory

The discussion thus far has cast doubt on the application of an impersonal conception of information. Although this conception is ubiquitous in the discussion of q-memory, however, it is not essential to it. Q-memory does not require questionable assumptions of information-links, deviant causal chains, and memory-transfers or transplants. The subject might just discover that although some of their own apparent memory-claims are strictly erroneous, they correspond in surprising detail with someone else's past, as recollected by them and confirmed by others and by collateral evidence. They might then come

[53] It is also worth noting that although the genotype is a constraint of sorts on the phenotype – on the actual appearance of the organism, whether it has blond hair for instance – environmental factors may intervene; someone may be fed something as a baby that permanently alters the colour of his hair.

to believe that they have q-memories of that person's past. This is the most conservative presentation of the phenomenon, involving everyday rather than science-fictional examples. It does not beg the question in favour of q-memory by presupposing an information-link in an impersonal, knowledge-independent sense, and so ought to be of broadest appeal. However, such is the appeal of materialism that some might find it more incredible that q-memory could occur, apparently without any causal story of how it came about.

The curious story of the Bournemouth holiday mentioned earlier illustrates this conservative version of q-memory. Reminiscing about my childhood, I seem to remember arriving by steam train at Bournemouth station for a summer holiday, carrying bucket and spade, and so forth. I recall the busy platform, the station nameboard, the noise and smell of the steam engine. However, my mother assures me that we never had a family holiday in Bournemouth; in any case, I learn, steam trains ceased operating that service before I was born. Now imagine that Derek Parfit overhears my reminiscences, which remind him uncannily of a childhood holiday he himself had in Bournemouth. When I draw, apparently from memory, a picture of the hotel where I thought I had stayed, Parfit comments on the remarkable likeness to his hotel. There was a large toy clock in the local park which, I seem to remember, greatly fascinated me; Parfit finds a strange similarity with events that he recalls. And so on. Proponents of q-memory will argue that, given this mounting congruence between my apparent memory-judgements and Parfit's past, a q-memory-link is plausible. (As noted earlier, there are limits on the prevalence of such links owing to their impact on personal identity.)

Although this approach is the most plausible basis for q-memory, it fails on a number of counts to vindicate it. The first reason is that personal memories form a holistic structure; they are not a collection of independent items pertaining to isolated events. If I came to believe that I was afflicted by more than a limited incidence of q-memories, I might find myself distrusting my childhood memories in their entirety – thus undermining my sense of self-identity. A second reason arises from the fact that, as Parfit concedes, a particular q-memory cannot yield knowledge of someone else's past unless one knows in general terms how it was caused. Thus the knowledge of the past which q-memories might yield is not immediate, and depends on the possession of extraneous information concerning their origins. The q-memory-judgement – in the Bournemouth example, the judgement concerning Parfit's past however it is formulated – must be an inference from "I seem to remember

arriving at Bournemouth" and "Memories purporting to concern a Bournemouth holiday in the 1960s in fact concern Parfit's past". Even if one were confident that the inferences in question yielded knowledge, its inferential status means that the q-memory-judgement cannot be memory-like.

This is an important objection. However, q-memory should be rejected not just because it yields only inferential knowledge of the past, but because it lacks most of the features essential to personal memory. Q-memory-judgements are made purely as a result of receiving indications after the event that it happened, and so do not constitute retained knowledge of the past. They are not spontaneous past-tense claims. Finally, the role of memory in generating self-identity is ignored. All that q-memory has in common with memory is that – on certain suppositions – it yields knowledge of the past, and subjects are inclined to confuse it with genuine memory. Q-memories are neither my memories, nor someone else's, for what makes a memory mine is that it concerns events that I have witnessed, and that I give utterance or other expression to it. This essentially dual nature of memory reflects the fact that memory, in all its varieties, involves information-retention and not information-transmission.

The grip of the transmission model on the philosophical and scientific psychological imagination means that it is common to regard information as transmitted within the same subject over time – hence the popularity of a causal analysis of memory. If one takes the concept of memory as retained knowledge seriously, however, the idea of q-memory can get no grip. As argued earlier, information has to be information for someone. Knowledge must be retained, that is, retained by someone; continued existence as a transferable trace is not sufficient for memory-based knowledge. And knowledge cannot be said to be retained if the identity of the person possessing it has changed – that is, if it has been transferred. This is not a trivial analytic claim; hence neither is it trivially analytic to say "My own witnessings or experiences are the only ones that I can remember".[54]

There is a further reason for not resting content with a critique of q-memory based on the merely inferential character of the knowledge it generates. This is because the resulting picture is espoused by those who accept the picture of a prior, determinate information-link, while also adhering to materialism concerning self-consciousness, and so constitutes

[54] Grice regarded this truth as analytic, while Shoemaker treats it as synthetic (Grice 1941, p. 344; Shoemaker 1984b, p. 22).

a way of trying to make IEM consistent with that kind of materialism.[55] An example is John Campbell's account of the guarantee of "not losing track of oneself", his laconic formulation of IEM. Campbell agrees with Evans and McDowell that memory-transplantation is conceivable, and that the result is not q-memory as ordinarily understood:

> So long as the first person has the memory ... At no point does he lose track. When the second person acquires the seeming memory, he simply acquires a false belief about what he himself did. At no point does he have to keep track of who is in question; at no point does he lose track of who is in question.[56]

Campbell does not question that the second person acquires a seeming memory, even though he allows that it does not result in a genuine memory-belief. Since B's knowledge that A swallowed a ball bearing, for example, depends on an inference, Campbell is right to conclude that there is no genuine counter-example to IEM here. But the description of the case as involving memory-transfer – without which it would not be a case of anything – makes it natural to say that B has misidentified the person who swallowed the ball bearing. The fundamental error is acceptance of the information-transfer model.

The present chapter has examined and rejected the reductionist critique of IEM arising from q-memory, defending the claim that first-person memory-judgements exhibit the immunity in question. In the next chapter, IEM is further defended as a substantive phenomenon and not merely an artefact of a method of description. There it is argued that memory and personal identity form a conceptual holism – so there is no vicious circularity, and no need for the postulation of q-memory to avoid it. The next chapter also broadens the range of cases of IEM, arguing that it is a feature of a range of future-tense as well as past-tense claims; subsequent chapters defend the claim that proprioceptive judgements also exhibit it.

---

[55] Cassam originally rejected IEM because he regarded it as inconsistent with that position; see Chapter 6.

[56] Campbell (1994), pp. 98–9. Cassam (1997a), p. 175, discusses the impossibility of pervasive q-memory.

3

Memory and Self-Consciousness (2): The Conceptual Holism of Memory and Personal Identity, and the Unity of Consciousness

This chapter continues the response to the dilemma previously discussed. If we define what it is for a subject to remember in a way that allows that they could be distinct from the subject whose experiences or deeds are being remembered, it seems that memory cannot constitute a sufficient condition for personal identity. But if we define what it is for a subject to remember in a way that rules out that possibility, it then seems that our definition must rest on a prior understanding of personal identity, and so the proposed account becomes circular.

Chapter 2 rejected the first horn of the dilemma by defending IEM against a reductionist critique. The present chapter shows that embracing the second horn of the dilemma need not result in vicious circularity, since there is a conceptual holism between memory and personal identity. It argues that IEM is a substantive phenomenon, not merely an artefact of a method of description, and that the charge of vicious circularity rests on the mistaken assumption that any account which inter-defines personal identity and memory must accord epistemic priority to one of the two concepts.

To say that personal memory exhibits IEM is to say that a sincere memory-claim in the continuous-verb form implies that either I o-ed, or else no one o-ed. It is this that undermines Butler and Reid's circularity objection to the memory criterion, and constitutes the criterion's qualified truth – that memory suffices for personal identity. That is, it is sufficient, though not necessary, for being some past X, that one remembers events witnessed by, and actions performed by, X, in an appropriately

personal or "inside" way. Indeed, the Butler-Reid objection to the memory criterion is a precursor of the objection that IEM is artefactual.

Perceiving the truth in the memory criterion means reconfiguring the traditional debate concerning personal identity, which requires foundational criteria. The change is achieved by recognising the *conceptual holism between memory, expectation, and personal identity*, which denies any priority between the unitary self and the unity of consciousness. The unity of consciousness is not conceived of as consciousness of an independent unitary thing or simple substance, nor is it interpreted as it is in Humean treatments, as the unity of the stream of subjective consciousness. Rather, it is interpreted in terms of IEM, in particular the IEM of memory-judgements – as the unity of present remembering with the remembered past.

3.1 The objection that the immunity to error is artefactual

The critique of IEM addressed in this chapter is both deeper and more elusive than that arising from information-garbling and q-memory. In responding to it, the significance of IEM for self-consciousness, and its connection with personal identity, becomes clearer. While conceding that the philosophical issue remains difficult to keep in focus, I argue that IEM is not an artefact of a method of description.[1]

It is the artefactuality objection that Gareth Evans seems to have in mind when he warns of "a danger inherent in all our reflections on self-consciousness", and argues that guarantees of re-identification of the subject over time may be "tautological". Evans explains the danger in a way that reflects a certain ambivalence. Simply in virtue of using personal pronouns, he writes

> We are building the subject's identity over time into the description of his situation. This may make it appear that he has an infallible knowledge of what is involved in this identity; but the appearance is nothing but an artefact of our way of describing the situation.

Nonetheless Evans believes that the method of description which uses personal pronouns, and takes the identity of the subject of thought for

[1] A reminiscence by Wittgenstein affords a neat comparison: "I said to [Frege] 'Don't you ever find *any* difficulty in your theory that numbers are objects?' He replied 'Sometimes I *seem* to see a difficulty – but then again I *don't* see it'" (reported to Geach; quoted in Wright 1983, p. xii).

granted, is "unavoidable".[2] In effect he poses a dilemma, comparable to our opening one: either it is presupposed that it is the same subject which at different times witnesses an event and then recalls it, in which case the claim that the later subject cannot misidentify the earlier one will be tautological. If this is not presupposed, however, it will then simply be a mistake to conclude from the fact that the different experiences are self-ascribed, that it is the same subject at different times which is experiencing and self-ascribing.[3]

My main concern is not the interpretation of Evans, but the objection to IEM arising from his considerations. However, Evans's remarks are puzzling because while he later defends IEM, it is difficult to interpret the second disjunct of his dilemma, which suggests that IEM is an artefact of one's description of the situation, as anything other than as a challenge to it. The claim that memory-judgements exhibit the guarantee, although not false, is regarded as true merely by convention.

His first example of the alleged tautology, it is true, does not relate to IEM, but to Anscombe's claim of no unnoticed substitution of the referent of "I". If "I" refers, she argues, it must have guaranteed reference to a evanescent Cartesian ego – a thinking immaterial subject freshly created with each use of "I". (Recall Lichtenberg's objection that Descartes was entitled only to "There is thinking", and not "I think"; Anscombe's transient ego is one step removed from Lichtenberg's complete elimination of a subject.) Anscombe concludes: "Even so, there is an assumption that something else does not surreptitiously take its place. Perhaps we should say: such an assumption is extremely safe for 'I', and it would be altogether an excess of scepticism to doubt it!"[4]

Evans is unimpressed with this guarantee:

> It is a simple tautology that, if it is correct to describe the situation [in terms of one and the same subject having thoughts at various times], the *self*-identifications are all identifications of the same self, and hence it cannot be a reason for anything.

He evidently assumes that the self-identifications to which Anscombe refers are entirely present-tense ones, made by the same subject at different times. But if her subject is portrayed as making past- or

---

[2] Evans (1982), p. 213.

[3] The passages where the dilemma is suggested are pp. 213–15.

[4] These remarks figure in her *reductio ad absurdum* argument against the position that "I" refers (Anscombe 1981b, p. 31).

future-tense self-identifications as well, Evans's objection is less persuasive.[5] On the latter interpretation, Anscombe is rejecting the wholesale substitution of subjects implied by pervasive q-memory: the suggestion that all of one's apparent memories originating from before a certain time, in fact originate in and concern someone else's past. It would indeed be "an excess of scepticism" for the subject to wonder whether this had happened.

It is Evans's second example that has a bearing on IEM.[6] He denies that the remembering subject, and the subject who witnessed or experienced the remembered events, must be identical:

> We might say: certainly, when time "t" comes, [the subject] will know whether or not the hypothesis that he expressed earlier by "I'll be in pain at t" was or was not correct, just [on the basis of] whether or not he is in pain at "t" ... It is not possible for the subject to have got hold of the wrong person at time "t" ... [since] as the case is described, there is a logical guarantee of adequacy. But this is, again, an artefact of our way of describing the situation ... The "method of verification" has a presupposition. Of course we must not say (using the ordinary vocabulary): it presupposes that the subject remains the same over time. But it presupposes that the subject who exists at "t" and "remembers" the hypothesis expressed earlier is the person who made the hypothesis, and hence is the person whom it concerns. And this is something of which he can have no genuine logical guarantee.

Evans adds that he does not want to deny criterionless self-ascription of anticipated properties, the other side of the coin to criterionless self-ascription of remembered properties. (For him, "criterionless" does not imply "immune to error"; errors may still be possible.) But he does want to deny that these features are a matter of a "logical guarantee of an identity assumption".[7]

In the long passage just quoted, Evans contrasts different ways of expressing a guarantee of self-identification: one of them is substantive; the other is in some way empty or tautological. The remark "It is not

---

[5] Evans (1982), p. 214. Hamilton (1991) granted the point too easily to Evans.

[6] Evidence that the dilemma concerns IEM also arises from his very interesting later remarks on the connection between IEM and the memory criterion for personal identity, of which more below.

[7] Evans (1982), pp. 214–15. The IEM of anticipated properties – of intention and expectation – is considered shortly.

possible for the subject to have got hold of the wrong person at time 't' ... " could refer to a past-tense claim made at t, or to a future-tense claim made before t about the subject-at-t. If the former, Evans is right to deny that the subject has infallible knowledge since, to reiterate, their judgement might turn out to be based on testimony and thus open to misidentification (though he nowhere mentions this possibility, and we have seen that he neglects it in his definition of q-memory). Moreover, it is wrong to say that "I'll be in pain at 't'" implies no logical guarantee of adequacy. If the justification that the subject would offer is based on personal memory – as in Evans's example, when the subject reiterates a hypothesis expressed earlier – then there is a guarantee, namely, IEM.

The second alternative opens up the question of future-tense IEM, which we examine before returning to the main discussion of IEM's alleged artefactuality.

3.2 Future-tense immunity to error

Future-tense IEM has been neglected in the literature, perhaps because it seems implausible. In fact, future-tense self-ascriptions do seem to exhibit an analogue of memory-based IEM. Those writers who touch on the issue disagree over what this analogue is, however.[8] Mill, in his Humean account of the self as a series of perceptions, cites expectation; Parfit discusses q-intention.[9] Describing the feature which, he believes, invalidates the series account of the self which he has earlier proposed, Mill writes that "The thread of consciousness which composes the mind's phaenomenal life, consists not only of present sensations, but likewise, in part, of memories and expectations ... ". Interestingly, he continues,

> In themselves, [these] are present feelings ... But they are attended with the peculiarity, that each of them involves a belief in more than its own present existence ... [a belief] that I myself formerly had, or that I myself, and no other, shall hereafter have, the sensations remembered or expected.

This is close to the phenomenon of IEM.

---

[8] Thus Evans is right to note that the criterionless self-ascription of anticipated properties is "simply the other side of the same coin as criterionless self-ascription of remembered properties" (Evans 1982, p. 215n).

[9] Parfit ref.; Mill (1979), pp. 193–4. Mill's account is discussed in Hamilton (1998a).

Self-ascriptions of intention are avowals, and as such are completely immune to error; all therefore exhibit IEM. [10] *Self-ascriptions of expectation* are not avowals, but some categories do exhibit IEM. Intention is central to the IEM of anticipated properties, but it is *expectations concerning my own future actions, experiences, or witnessings* which mirror the case of memory. As in the case of memory, IEM applies only to *personal* as opposed to *factual expectation*. "I expect to meet Mrs Thatcher, but maybe it will not be myself who meets her but someone else" makes sense only if its grounds do not involve intention. Otherwise the judgement is IEM. That is, while I can doubt that I will meet Mrs Thatcher – "I expect to meet her, unless my train is much delayed" – if the grounds of the judgement remain constant, I could not retreat to wondering whether it will be myself or someone else who will meet her. Contrast where I am one of a group of admirers, who are told that Mrs Thatcher wishes to meet one of us, and on the basis of other evidence I infer that it will be myself, but then learn that my informant is confused about whether it is a philosopher or a psychologist that she wants to meet. It would now make sense for me to say "I expect to meet her, but it is possible that it will not be myself but someone else who meets her". It may be that the content of my intention has a bearing also. If it is the same as the content of the expectation – that is, to meet Thatcher – it is probable that the corresponding expectation-judgement is IEM. But if the content is different – for example, to make sure that some philosophers, hopefully including myself, meet Thatcher – the corresponding expectation-judgement is not IEM.

IEM is exhibited also by judgements of expectation based not on intention but on personal memory. "On past evidence, I won't book up my summer holiday until the last minute", is IEM when based on my own recollection, rather than learning about my past actions from someone else's testimony, as an amnesiac stroke victim might do. One might say "If past evidence is any guide, it will be myself who gets to meet Mrs Thatcher", where the past evidence is my own recollection; but here past evidence is a guide as to *who* will meet her. If the judgement "I expect I'll see you there tomorrow" is based on my intention to be there plus my belief, itself based on evidence, that you will be there, it is not. If it is based on my prediction that I will be there, arising perhaps from recollection of my past habits, it is not IEM. One may therefore conclude that expectation exhibits IEM when based either on intention or on personal memory, or both, possibly combined with further evidence. It follows that in Evans's example, "I'll be in pain at t" is IEM

[10] Avowals of intention are the subject of Hamilton (2008).

if I intend to harm myself at t. It is not IEM where I am told that I will have an operation which would result in my being at pain in t, since judgements based on testimony do not exhibit IEM – the person who told me could be confusing me with another patient.

3.3 The objection that the immunity to error is artefactual...

Whether "It is not possible for the subject to have got hold of the wrong person at time 't'..." refers to a past-tense claim made at t, or to a future-tense claim made before t, the phenomenon that Evans is describing must, I think, be IEM, and not the appearance of infallible knowledge of one's own identity. But why would he object to the phenomenon here, while elsewhere defending it? There are two possibilities. First, although he apparently regards IEM as fundamental to self-consciousness, he also regards it as consistent with weak q-memory. Commitment to IEM is in tension with his materialist identification of the subject as physical entity, and his conviction that in using "I", the subject is tracking a physical object – that is, to his commitment to a traditional, robust conception of personal identity which makes the unity of consciousness subservient to a materialist unitary self. [11] This is why he endorses only the qualified claim that self-consciousness involves capacities that *normally* exhibit IEM – abnormally they may at least "yield information" about a person distinct from the subject.[12]

A second explanation of Evans's remarks, not inconsistent with the first, is that there really is an issue about the artefactuality of IEM. The challenge that IEM is artefactual – a product of a method of description – is *conventionalist*.[13] Conventionalism about persons, like Parfitian

[11] Though this is the dominant position in Evans's discussion, it contrasts with his account of memory as a retentive capacity: "...so far as the 'I'-Idea is concerned, the later dispositions to judge flow out of the earlier [ones], without the need for any *skill* or *care* (not to lose track of something) on the part of the subject" (Evans 1982, p. 237).

[12] The issue is well presented in O'Brien (1995a), to which I am indebted.

[13] There is a parallel issue concerning proprioception-based IEM and the individuation of subjects, in which the Evans-style warning would be: "It is a simple tautology that, if it is correct to describe the situation [in terms of distinct subjects having thoughts at the same time], the *self*-identifications are all identifications of distinct selves, and hence it cannot be a reason for anything". The concern about guarantees over time parallels those at one and the same time. Chapter 5 offers a response to that dilemma, analogous to the one presented here.

reductionism, denies that the identity of the subject of thought has to be taken for granted; unlike reductionism, however, it treats the IEM claim as empty rather than false. This contrast is evident from the aforementioned ambivalence in Evans's presentation of the artefactual objection. He claims that the identification guarantee is an artefact of the method of description which uses personal pronouns – yet concedes that this method, which takes the identity of the subject of thought for granted, is "unavoidable". Since it makes sense to talk of a linguistic artefact only if one has a genuine choice concerning the means of description, Evans's concession shows an ambivalent attitude. This attitude is shared with John Campbell, who claims that the identificational guarantee is in some sense a product of one's description of the situation, though "it is by no means a superficial matter that this is the right way to describe [it]".

Both writers, it seems, see the attraction of treating IEM as a priori – the view defended in the present volume – even while affirming its conventional or artefactual status. On Campbell's view, the guarantee of "not losing track of oneself" – his laconic formulation of IEM – is not a priori, but depends on the "background contingency" that we are creatures whose psychological lives exhibit an integrity over extended periods. Creatures who lack this integrity, and fuse or divide, would have no equivalent of the first person, nor any beliefs or desires, Campbell argues.[14] I would argue against Campbell that such creatures are not self-conscious – since, as suggested in Chapter 1, for self-conscious creatures, the "method of description" using "I" is unavoidable, and not merely one possible conceptual scheme among others.

Now we can recognise that a further condition is required: that the self-reference is made by a persisting subject. It may be a contingent matter that self-conscious creatures exist, but if they do, it is a priori that their thought involves the first –person – and, as Evans says, use of the first person takes the subject's identity for granted. IEM expresses that identity, for, as I will argue, it is the substantive phenomenon underlying the unity of consciousness of remembered and remembering subject – the only non-tautological or non-trivial guarantee concerning "I". IEM-exhibiting ways of knowing necessarily yield knowledge of, and information from, the subject. This is not quite the "logical guarantee of identity" which Evans rejects, since it does not rest on a claim

[14] Campbell (1994), pp. 95–8. The analogy which Campbell uses to illustrate his claim was argued to be unconvincing in Chapter 2.

of infallible knowledge of one's own identity, but it requires that IEM is neither false nor artificial.

The remainder of this chapter responds to the conventionalist challenge by showing how memory- and intention-based IEM rests on a conceptual holism between memory, expectation, and personal identity. This response shows how IEM expresses the qualified truth in the memory criterion, and dissolves Evans's dilemma by reconciling the unity of consciousness and unitary self standpoints. In particular, the chapter shows that the identity question, "What are the criteria for personal identity?", is no more fundamental than the unity question, "What is it that unites a person's present experiences with their past ones?" A unity of consciousness standpoint, in the qualified form defended here, is consistent with holding that what unifies a single set of experiences is simply that they are all had by one person, that there is no further fact concerning personal identity. That is, since the question "What is it that unites a person's present experiences with their past ones?" presupposes that it is the same person, there is nothing further to say after answering the identity question. This does not, however, imply that the identity question is more fundamental than the unity question.

3.4 IEM, personal identity, and conceptual holisms

The "method of description" which takes the identity of the subject of thought for granted is unavoidable, I believe, because memory and personal identity constitute a conceptual holism. Proponents of the artefactuality objection share Butler and Reid's assumption of a unitary self. The connection between IEM, self-consciousness, and personal identity, implicit in the preceding discussion, lies in the fact that the Butler-Reid objection to the memory criterion is a precursor of the objection that IEM is artefactual.[15] Both objections assume that the concept of personal identity has epistemic priority over that of memory – the claim denied by the conceptual holism of memory and personal identity. I embrace the second horn of Evans's dilemma, "presupposing" that it is the same subject which at different times witnesses an event and then recalls it, while rejecting Evans's implication that the result is a tautological guarantee.

IEM has important consequences for the personal identity debate because, as I will argue, it undermines the circularity objection to the

[15] Butler (1896) I, p. 385; Reid (1967), p. 352, EIP [full ref].

memory criterion. This crucial link is noted by Evans when, after arguing that memory-judgements are IEM, he continues:

> [IEM] has a connection with the idea that memory suffices for personal identity. My judgement (or seeming judgement) "I was F", made on the basis of apparent memory, can go wrong only in certain ways. Either there is no past state of affairs of which the apparent memory constitutes information (no one was – relevantly – F); or the object involved in the state of affairs was indeed myself; or finally, *I* do not exist.[16]

What is the "connection" between memory-based IEM and the memory criterion of personal identity?

In speaking merely of a "connection", Evans draws back from saying that IEM *implies* that memory is sufficient for personal identity. Note that the connection is not with the memory criterion as such, which says that memory is both sufficient *and* necessary for personal identity – that remembering events witnessed by, and actions performed by, X, in an appropriately personal or "inside" way, is necessary and sufficient for being identical with that past X. Only the sufficiency claim is plausible; the necessity claim is inconsistent with maintaining one's identity with early childhood, before memories appear. However, I will argue that IEM constitutes the memory criterion's qualified truth: that memory is sufficient for personal identity in the sense that when "I was F" is made on the basis of personal memory, my grounds for believing that someone was F are grounds for believing that it was myself who was. To reiterate the Introduction, a cloud of metaphysics of personal identity is condensed into a drop of grammar concerning the epistemology of self-consciousness.

This process involves rejecting the traditional search for foundational criteria of personal identity. When treated as a solution to the traditional problem, the memory criterion is indeed decisively refuted by the Butler-Reid objection, according to which the memory criterion confounds "personal identity … with the evidence which we have of our

[16] Evans (1982), pp. 246 and footnote; Evans is outlining the range of options – "the object involved in the state of affairs was indeed myself" is clearly not a way of "going wrong". The final possibility involves the brain in a vat scenario, discussed in Chapter 5. McDowell defends a view close to Evans's in his (1997).

personal identity", and thus involves vicious circularity.[17] Their view is the consensus, as shown in Shoemaker and Parfit's reductionist interpretation of psychological criteria, which invokes q-memory to avoid the alleged circularity.[18] Properly understood, however, IEM reveals an alternative approach which undermines the Butler-Reid objection by showing that memory and personal identity comprise a holism of interdefined concepts.

Interestingly, there is a sketch of this holistic alternative in J. S. Mill's little-known *Notes to "The Analysis of the Human Mind"* by his father, James Mill. Here Mill writes,

> The phenomenon of Self and that of Memory are merely two sides of the same fact ... We may, as psychologists, set out from either of them, and refer the other to it ... But it is hardly allowable to do both. At least it must be said, that by doing so we explain neither.[19]

Mill implicitly rejects Butler and Reid's complaint – of which he would have been aware – that memory presupposes personal identity and so cannot constitute the criterion for it. Unfortunately, his discussion has an unresolved tension between first- and third-person criteria for personal identity. He criticises Locke for ignoring third-person (bodily) criteria, yet endorses the latter's analysis of first-person (psychological) criteria: "My personal identity consists in my being the same Ego who did, or who felt, some specific fact recalled to me by memory". Mill does not explain why two different sets of criteria supply the same answer to questions of personal identity, and he may not even recognise that it is the same question which they address. Nonetheless, his suggested conceptual holism between memory, expectation, and personal identity offers a novel response to the Butler-Reid objection, pointing to a circularity which, unlike most writers, Mill does not find vicious.

3.5 The circularity objection, no-priority explanation, and the holism of memory and personal identity

We now consider *Euthyphro*-type paradoxes in order to undermine the standard assumption of circularity objections, that one of the pair of

[17] Reid (1967), p. 352. Although it is debatable whether Locke really subscribed to a memory criterion, he is commonly taken to have done so. The interpretative question is well discussed in Southgate (1999).
[18] See for instance Parfit (1971), p. 15.
[19] Mill (1989), p. 138; pp. 212–13.

concepts in question must be more fundamental – hence the requirement of foundational criteria for personal identity. Plato's paradoxes may have concerned causal or explanatory, rather than epistemic, priority. Here they are used, perhaps anachronistically, to illuminate the *holistic no-priority proposal* that I am making. According to the original Platonic paradox, either God's will or goodness is fundamental.[20] But recognition of conceptual holism yields a distinctive *no-priority resolution* – converting the dilemma, effectively, into a trilemma. On my reinterpretation of the original *Euthyphro* paradox – whether God wills the good because it is good, or the good is good because God wills it – the neglected alternative claims a conceptual holism between God and goodness. Thus the traditional objection to the divine command interpretation, that God might will something morally outrageous by secular standards, could not arise. It might be argued that this third alternative really just amounts to the first, constitutive disjunct of the dilemma – that which Wittgenstein, in characteristic nonrationalist vein, regarded as more profound – since proponents of the latter agree that in learning about goodness one is learning about God.[21] But it does appear that divine command theories, in regarding God's command as antecedent to morality, make the connection between God and goodness too contingent.[22]

The unclarity of the divine command standpoint results at least partly from the obscurity of religious concepts. Two rather more clear examples are found in aesthetics. The first concerns the definition of art, and arises from Dickie's claim that "X is an artwork if and only if it is accepted by the artworld as a candidate for appreciation".[23] The biconditional apparently leaves open the question of whether X is an artwork because the art world declares it is, or whether they declare it so, because it is antecedently an artwork. (The sense of "art" in question is the modern conception of fine art.) However, from a conceptual holist standpoint, the truth expressed in the biconditional is that aesthetic

---

[20] Recent debate on the paradox is discussed in O'Sullivan (2006).

[21] See Wisdo (1987).

[22] Ontologically, if the good is identified with whatever God's nature causes him to approve of, then there are not two things, but one, and nothing is prior to itself. But still there could be, and perhaps would have to be, an explanatory priority of the kind physicists suppose the physical has over the mental: theists explain the goodness of something in terms of what a being who represented perfect love would want.

[23] Dickie (1974).

appreciation in institutional settings such as art galleries and concert halls arises together with the modern concept of art.[24]

A second aesthetic example concerns the basis and objectivity of aesthetic judgement. Hume argues that "the joint verdict of [the true judges] is the standard of taste and beauty", which is ambiguous, falling between "The critics' verdict determines aesthetic merit" – an application of Crispin Wright's "best opinion" schema, that the judgement of critics constitutes or is a criterion of beauty – and "Their verdict reflects (antecedent) aesthetic merit".[25] My suggestion is that there is a third option, according to which the truth in the biconditional "X has aesthetic value if and only if the true critics judge that it has" lies in a conceptual holism between aesthetic judgement and critical practice, for instance, that any genuine aesthetic judgement can be justified by critical expertise or experience.[26] In acquiring the concept of aesthetic value, one acquires the concept of critical judgement, and vice versa.

This standpoint has recently received some attention from David Wiggins, whose account of property-response pairs involves a "subjectivism of subjects and properties *mutually* adjusted".[27] John McDowell also defends a no-priority alternative to the opposition of intuitionist realism, and the metaphysics of projectivism, according to which neither moral sentiments nor moral properties are basic, rather they are mutually explanatory. He criticises Blackburn's opposition between realism,

---

[24] The example is discussed in Hamilton (2007a), and pursued in Hamilton, *The Claim of Art* (forthcoming 2014).

[25] Hume, in "Of The Standard of Taste". In Wright's term, these judgements are extension-determining and extension-reflecting respectively (Wright 1992, pp. 108ff).

[26] It is not clear whether this claim amounts to a distinctive resolution of the *Euthyphro* paradoxes, or just a distinctive philosophical position with no distinctive resolution. Only a constitutive interpretation of the biconditional – that which refers to "best opinion" – could be consistent with a conceptual holism; a reflective interpretation is excluded. If it is conceded that one cannot have artworks without an institutional setting, then an object's status as artwork cannot be prior to the art world's acceptance. One may allow that in general artworks require an art world, but deny that a particular object has to be validated by the art world for it to count as an artwork. But in that case the object would need intrinsic features independent of the existence of the art world, which undermines the claim that the latter is required. The question is whether the holistic standpoint asserts a parity between critical judgement and aesthetic value, for instance, instead of prioritising critical judgement in the manner of the constitutive account.

[27] Wiggins (1987), p. 199.

which holds that the moral features of things are the parents of our sentiments, and Humeanism, which holds that they are the children; McDowell treats values and sentiments as siblings, maintaining that this no-priority view is consistent with a standard of correctness in moral judgement.[28] In contrast to projectivism, which tries to show that by non-moral standards some sensibilities are superior, the no-priority view argues that such reasons do not require a standpoint external to ethics, rather we make use of our other moral judgements.

The conceptual holist position has been canvassed in various areas of philosophical debate, if not under that name, and without generating much enthusiasm. For instance, Brandom considers the "no-priority" view that explaining representational properties of beliefs involves considering their inferential role, and vice versa, without endorsing it or explaining his reasons. Peacocke examines the "no-priority" view that being red and looking red have to be introduced simultaneously. Finally, Brewer rejects a "no-priority" view concerning the relation between "the perceptible properties of physical objects and the various appearance properties that show up in our perception of them".[29] The no-priority position advocated here is more general, however. Analytic philosophy, not surprisingly given its name, is concerned to analyse. But the picture of conceptual holism defended here is itself a form of analysis – more plausible and sophisticated than the received foundationalist picture.

Returning to the circularity objection to the memory criterion, we can see that it shares the assumption of conceptual priority found in discussions of Euthyphro-type paradoxes. The objection holds that personal identity is more fundamental than memory or expectation; the latter, for simplicity, I will omit, and just refer to memory. Thus Butler and Reid seem to believe that personal identity must be explained independently of memory, which Locke fails to do. Reid claims that he confuses "personal identity … with the evidence which we have of our personal identity", while Butler writes: "One should really think it self-evident that consciousness of personal identity *presupposes*, and therefore cannot *constitute* personal identity, any more than knowledge … can constitute

---

[28] McDowell (2001), p. 159. The no-priority view "opens up the possibility that it might be respectable to *use* the apparently world-describing conceptual resources with which we articulate our responses, in earning truth in one of the relevant areas" (op. cit. p. 160).

[29] Brandom (2000), p. 205n6; Peacocke (1997); Brewer (2011).

truth, which it presupposes". Reid is scathing about what he regards as the idealist implications of the memory criterion:

> to say that my remembrance that I did such a thing, or my consciousness, makes me the person who did it, is ... an absurdity too gross to be entertained ... for it is to attribute to memory or consciousness, a strange magical power of producing its object, though that object must have existed before the memory or consciousness that produced it.[30]

He claims a parallel between "I would not be said to be remembering if it had not been me that did it", and "I would not be said to know that p, if it were not the case that p". Reid seems to object that Locke inflates an analytic truth – the factive status of remembering – into a species of metaphysical nonsense.[31]

This line of objection would be justified if one had to assume that either memory or personal identity must be conceptually prior. The claim that to have memory-experience of an action as having been performed by myself, or an event as having been witnessed by myself, is necessary and sufficient for it having been performed or witnessed by myself, is indeed absurdly idealist, as Reid recognises.[32] (It contrasts with the more plausible idealism – an idealism of ownership only – which underlies Locke's account of ownership of body parts, discussed in Chapter 4). It does not follow that memory is mere evidence of personal identity, however; nor

---

[30] Butler, "Of Personal Identity", in his (1975), p. 100; Reid (2002), p. 277. Reid did not use the term "idealist" in this sense, since for him it referred simply to subscription to the Theory of Ideas. Ayers (1991), p. 269, argues that Leibniz advanced a similar objection to Reid's.

[31] Strictly speaking, factivity is the property of licensing only content clauses that represent claims assumed to be true, so it is "I remember that p" that is factive. It should be noted that Locke sidesteps the objection by distinguishing persons from thinking substances; sameness of consciousness is a relation between thinking substances rather than persons. Clearly he then faces problems in sustaining the distinction (Noonan 1989, pp. 69ff.).

[32] Child (2006) attributes to Wittgenstein something like this position, what he calls the "anti-realist" view that our judgements about past intentional states, namely, personal memories, are constitutive of the truths they seem to report, a position he opposes to realist common sense. He points to a series of remarks that, he says, suggest the anti-realist view; for instance that there was a past thing meant, intended, or wanted, whose identity is constitutively determined by the subject's retrospective judgement. He rightly finds the latter view implausible, but to attribute it to Wittgenstein is, I think, tendentious.

is memory merely consciousness of some underlying identity. The fact that a sincere memory-claim in the continuous-verb form implies that either I o-ed, or else no one o-ed – that personal memory exhibits IEM – undermines the circularity objection.

To explain why this is so, we turn to Noonan's pellucid defence of the objection. An example of a viciously circular explanation, he writes, is one which defines bachelor as an unmarried man, and then goes on to define being unmarried as being an adult bachelor or spinster (to mention a term now politically incorrect). No one with their wits about them would do this, Noonan argues, nor would they define personal identity in terms of memory, and then go on to define memory in terms of personal identity. (It this lamentable reasoning that concerns J. S. Mill in the quotation above – that "We may … set out from either [the Self or Memory], and refer the other to it … But it is hardly allowable to do both".) So this cannot be what the circularity objection is driving at, nor, Noonan continues, can the objection be that circularity arises from the fact that memory implies or logically entails personal identity, since such an implication must occur for any definition to be correct.

What the circularity objection affirms, Noonan argues, is the epistemological priority of personal identity over memory – that in order to distinguish genuinely remembering from seeming to remember, one must already have the concept of personal identity.

> To establish that someone not only *thinks* that he remembers … but actually *does* remember doing or experiencing something, we have first to establish that he, that very same person, did indeed do or experience it … The point is that the *conclusive verification* of the proposition that someone remembers doing so-and-so would have to involve checking that he did indeed do it … But this means that someone must already have the concept of personal identity if he is to have the concept of memory.[33]

Noonan's presentation of the circularity objection as a "priority of explanation" objection seems to me absolutely correct. Its challenge is that concept X is being used to explain concept Y, when in fact it presupposes concept X, and so the order of explanation must be incorrect. Where I disagree with Noonan is over his assumption that personal identity is indeed conceptually prior.

---

[33] Noonan (1989) pp. 171–2.

In fact, personal identity is prior only from a third-person viewpoint. When someone has been observed o-ing, and X claims to have o-ed, the question can indeed arise whether X was that person. The first-person case, in contrast, exhibits IEM: when I judge on the basis of memory that I o-ed, it makes no sense for me to attempt to verify that it was myself as opposed to someone else who o-ed. There is no question in this case of having to establish personal identity. It would be absurd for me to say, "To establish that I not only *think* that I remember o-ing, but actually *do* remember o-ing, I have first to establish that it was myself rather than someone else who o-ed". IEM precisely says that I do not have to know that it was the same person in order to know that I did it; it makes no sense to suppose that I could know that someone o-ed without knowing that that someone was myself.

We can now see that Evans's suggestion of a connection between IEM, and the sufficiency of memory for personal identity, is the qualified truth in the memory criterion, as follows:

> X can legitimately judge, on the basis of remembering or apparently remembering o-ing, that they o-ed *if and only if* either X = the person who o-ed, or else no one o-ed.[34]

When one recognises that this phenomenon does not amount simply to the factive nature of remembering, Noonan's claim that I do not have to have the concept of personal identity before I can remember is undermined. We see instead that the concepts arise together. To say that personal identity is epistemologically prior is, I would argue, already to endorse a substantivist or unitary self-account of that concept. (As we will see later, analogous considerations operate in the case of proprioception and bodily identity.[35])

Non-holistic analysis of a concept must show how it is acquired through epistemically prior concepts. Holistic analysis also must offer a reflective

---

[34] With appropriate provisos concerning apparent memory, as discussed earlier.

[35] Chapter 5 undermines the circularity objection to a self-conscious knowledge criterion of bodily identity on analogous grounds to those offered here. The truth in the criterion for bodily identity is the IEM guarantee for proprioception:

> X can judge, on the basis of proprioception, that their arm is raised
> *if and only if* either X = the person whose arm is raised, or else no one's arm is raised.

In the first person, when my judgement is based on proprioception, I do not have to verify that a body part is mine as opposed to someone else's, just as in the memory case I do not have to verify that it is myself as opposed to someone else who o-ed.

philosophical reconstruction, illustrated by empirical investigation, of how concepts are (jointly) introduced – for instance, by example rather than by verbal definition. In the case of memory and personal identity, a child learns the use of the past tense by being immersed in the past-tense usages of others: the memory- and testimonial-judgements of its elders. It would not follow that the child acquires the concept of personal identity before that of memory – that they observe persisting individuals around them, and then learn that these individuals have a certain kind of access to their past. Rather, to acquire the concept of a persisting human individual, just is to acquire the concept of an individual with that kind of access to their past. One might say that the child is aware of persisting objects, but that the idea of persisting individuals arises with that of personal memory among other concepts. Recognising another as a persisting subject involves knowing that they have a past. I know about someone else's past when I recall past actions of theirs, or when I accept their memory-claims concerning those actions. That is how we learn about persisting subjects or persons, in contrast to persisting objects.

3.6 The unity of consciousness versus the unitary self: Kantian origins of the dialectic

Finally we must show that, contrary to current consensus, there can be a *non-reductionist* standpoint of the unity of consciousness. This final section shows how Kant advocates such a position. Concerning self-consciousness and personal identity, his advance on Hume lay in understanding that the unity of consciousness is not contingent – a causal condition of experience – but transcendentally necessary.

We should begin by stressing the links between conceptual holism and the unity of consciousness. The claim of a conceptual holism of memory, expectation, and personal identity helps to explain how the IEM guarantee arises. It contrasts with non-holistic explanations, such as the explanation of the IEM guarantee as arising from the nature of memory alone, which are more plausible in the case of restrictions on sensory knowledge. The sense of smell gives knowledge only of smells, though smells are not constituted by the ways in which they are known; perceptual faculties give knowledge only of the egocentric environment.[36] In some sense it is the nature of memory which yields the guarantee, since memory can concern only that which the remembering

[36] This is one assumption behind the "sole-object view" in Martin (1995).

subject has experienced or witnessed. But to assume that the guarantee has to arise either from the nature of the subject, or from the nature of memory (or analogously with proprioception), is to beg the question against the conceptual holism of personal memory and personal identity, or proprioception and bodily identity.

To say that there is a conceptual holism between memory, expectation, and personal identity is to say that the unitary self does not have conceptual precedence over the unity of consciousness – that the latter does not consist simply in consciousness of an independent unitary thing or simple substance. As Michael Ayers writes, "after Hume and Kant, both deeply influenced by Locke, it can hardly be considered beyond all question that a unitary self is prior, whether epistemologically or ontologically, to the unity of consciousness".[37] Ayers goes on to suggest that Butler and Reid's circularity objection assumed that the unity of consciousness consists in consciousness of an independent unitary thing or simple substance – which is what Locke's memory criterion questions.

Ayers's last claim seems absolutely right. However, it is misleading of him to bracket together Hume and Kant, with their radically opposed conceptions of the unity of consciousness. To claim priority for the unity of consciousness over the unitary self need not imply Humean or Parfitian reductionism. It is true that it is generally assumed that a unity of consciousness approach involves reductionism or constructivism about the self, and a contingent connection between experiences – the approach adopted by Hume, by phenomenalists such as Mill, Mach, and Carnap, and most recently, perhaps, by Parfit. This volume rejects a reductionist interpretation of the unity of consciousness in favour of a more Kantian interpretation, which regards the connection between experiences as a necessary one involving IEM.

These distinct interpretations of the unity of consciousness are often left unclear. Ayers refers laconically to a "unity at the phenomenal level", and quotes Leibniz's reference to "an identity which is apparent to the person concerned – one who senses himself to be the same".[38] The confusion between different interpretations of the unity results from an ambiguity between *the phenomenal unity of the stream of consciousness* – the feeling that one is meant to have that one's successive or simultaneous experiences, of a beautiful sunset, a pang of hunger, and a moment

[37] Ayers (1991), vol. ii,p. 270.
[38] Ayers (1991) vol. ii, pp. 205, 270.

of melancholy reflection, are necessarily connected – and *the necessary unity of the remembering or anticipating subject*, namely, the apparent fact that I am the same person as the one who had or will have certain experiences, or who performed or will perform certain actions.[39] The latter sense is my primary concern. The necessary connection is between the subject's recollection and the recollected experience – or the subject's expectation and the anticipated experience – and consists in the fact that either the subject o-ed or will o, or else no one o-ed or will o. This connection is captured by the phenomenon of IEM. Far from implying a reductionist analysis of persons, the resulting conception of the unity of consciousness is opposed to it. In spite of A. J. Ayer's claims, the unity of consciousness in my sense has no affinity with reductionism.[40]

The anti-reductionist interpretation of the unity of consciousness originates in Kant. Indeed, as already indicated, both sides of the self-consciousness dialectic – the unity of consciousness and the unitary self approaches – originate in his work. Their modern manifestations each offers a distinctive conception of what Kant's advance on Hume consisted in. Although Kant's account of self-consciousness was not consciously directed at Hume, one grasps his notion of the transcendental unity of apperception – as one does more obviously with his accounts of causality and substance – by considering it as a response to his precursor.[41] The dialectic is a familiar one. Hume argues that the idea of an identical self arises – naturally and unavoidably but without the possibility of philosophical legitimation – from experience. Kant agrees that no *empirical* legitimation of the concept is possible; no unitary self could be presented or inferred from experience. But he does not agree that there can be no *philosophical* legitimation; there is a transcendental legitimation of the unified subject as presupposition of experience.

Michael Ayers, as we saw earlier, stresses a unity of consciousness over the unitary self interpretation of Kant. He focuses on the Paralogisms, where Kant argued that the unity of consciousness must not be confused with consciousness of a Cartesian unity, and the earlier sections of the Transcendental Deduction, where apperceptive self-consciousness is portrayed as a necessary unity which arises as a condition of the

[39] The former has been much discussed, by writers from William James to Michael Tye (2004).

[40] Ayer (1963), p. 116.

[41] Guyer (2008), pp. 150–60, is one of the few writers to consider Kant's discussion as a response to Hume, but his account is rather compressed and does not always highlight the main issues effectively.

combination or synthesis of representations.[42] On this interpretation, Kant's advance on Hume consists in the realisation that the unity of consciousness is not a merely causal condition of experience, but is transcendentally necessary.

The second and contrasting aspect of Kant's approach to self-consciousness, downplayed by Ayers, accords priority to the unitary self. Interpreters such as Strawson, who favour this approach, find its origins in the claim of the Refutation of Idealism – added to the second edition of the *Critique of Pure Reason* – that inner experience is possible only on the assumption of outer experience.[43] The unitary self approach inspires the Strawsonian requirement mentioned earlier – that the self-conscious subject must conceive of itself as an embodied entity, a person among persons, able to locate itself in an objective world. On this interpretation, Kant's major advance on Hume consists in acknowledging the importance of the subject's embodiment.[44] Although Strawson and his successors have recognised and even emphasised considerations that support a Kantian unity of consciousness standpoint, their commitment remains to the unitary self. Illustrations of this are Cassam's advocacy of "materialism concerning self-consciousness", and A. J. Ayer's view that "a person's ownership of states of consciousness consists in their standing in a special causal relation to the body by which he is identified", which assumes that one can first identify the body and then assign the experiences.[45] When contemporary writers consider whether the unity of consciousness might be conceptually prior to a unitary self, they treat it only in its Parfitian, reductionist version.

Several writers rather perversely interpret Kant's own treatment as reductionist. For instance, Gareth Evans seems to defend this interpretation when he suggests that the danger of tautology concerning re-identification of the subject over time is what Kant may have had in mind when he spoke in the Third Paralogism of the "logical identity of the 'I'".[46] We now show that when Kant gives priority to the unity of consciousness over the unitary self, reductionism is not the result. The gloss that Evans and others put on the Kantian interpretation – that

---

[42] Ayers (1991), vol. ii, p. 270.
[43] Kant (1929), for instance (B275).
[44] Kant's misinterpretation of Descartes's account of embodiment and self-consciousness is discussed in Chapter 5.
[45] Cassam (1997a); Ayer (1963), p. 116.
[46] Evans (1982), p. 214n.

the subject must assume its own identity, even though there may in fact be no such identity – is mistaken. Kant does not suggest a distinction between merely considering oneself as a persisting entity, and actually being one. The Paralogisms perhaps state more clearly than the Transcendental Deduction that a unified subject is a presupposition of experience. In the Paralogisms, Kant argues that the "necessary unity of the subject" cannot be expressed in an analytic proposition or shown to be synthetic a priori from mere concepts.[47] Nor could this unity, "as a condition of every thought", be derived from experience. Rather, it is a necessary condition of experience: "This proposition ['I think'] is ... but the form of apperception, which belongs to and precedes every experience; and as such it must always be taken only in relation to some possible knowledge, as a *merely subjective condition* of that knowledge".[48]

This apperception or self-consciousness condition may seem close to a tautology. The description "necessary condition of experience", it may be argued, is elliptical for "necessary condition of experience had by a unified subject"; Kant appears to be arguing, otiosely, that experience requires a unified subject. It is not otiose, however, when one considers the accounts of Hume, and twentieth-century reductionists influenced by him – so again, Kant may be interpreted as responding from a Humean standpoint. He has a distinctive account of what unifies the subject, in terms of their self-ascription of experience. A unified subject is one that is able to judge that it is unified. Kant claims that for a representation to be mine I must be able to self-ascribe it, to judge that it is mine. He also defends a second claim which may be termed the *self-ascription requirement*: a self-conscious subject must be capable of self-ascribing representations, and it must be possible for the "I think" to accompany all my representations.[49] Kant writes that "it is only because I ascribe all perceptions to one consciousness (original apperception) that I can say of all perceptions that I am conscious of them". He continues: "The thought that the representations given in intuition one and all belong to me, is therefore equivalent to the thought that I unite them in one self-consciousness, or can at least so unite them". Apperceptive self-consciousness, for Kant, thus involves a capacity to

[47] Collins however argues that Kant does present the necessary unity of the subject as an analytic truth (Collins 1999, p. 122).
[48] Kant (1929), A354.
[49] Kant (1929), B131. The label comes from Cassam (1997a), pp. 118–19.

self-ascribe representations; it does not simply amount to an awareness of an object, the self, as Descartes assumed.[50] It involves *reflexivity*, the ability to conceive of oneself *as subject as object* – a modern conception of self-consciousness.

The self-ascription requirement implies that the subject thinks of itself as persisting, and grasps its own identity over time – a presupposition that Kant repeatedly emphasises in the Transcendental Deduction:

> We are conscious *a priori* of the complete identity of the self in respect of all representations which can ever belong to our knowledge, as being a necessary condition of the possibility of all representations. (A116)[51]

Commentators are divided between *substantive* and *formal readings* of this "consciousness" of self-identity. The substantive reading claims that the subject has actual knowledge of its identity, while the formal reading claims that it is necessary for experience that the subject presuppose its own identity. The former is advocated by Henrich:

> in every actual case of self-consciousness we know that every thought can be accompanied by the consciousness "I think ... In precisely this sense we have *a priori* knowledge of our identity, in so far as this consists in the sameness of the subject referred to by "I" in indefinitely many instances of "I think"-consciousness.

This interpretation cannot be correct, however. Such knowledge would have to involve a Cartesian awareness of one's numerical identity – one of the claims of Rational Psychology which Kant is concerned to undermine.[52]

---

[50] Kant (1929), A122, B134. Note also: "... the empirical consciousness, which accompanies different representations, is in itself diverse and without relation to the identity of the subject [as Hume recognised]. That relation comes about, not simply through my accompanying each representation with consciousness [as Descartes claimed], but only in so far as I *conjoin* one representation with another, and am conscious of the synthesis of them. Only in so far, therefore, as I can unite a manifold of given representations in *one consciousness*, is it possible for me to represent to myself the *identity of the consciousness in [i.e. throughout] these representations*" (B133, my parenthetical comments).

[51] There are similar comments at A113, A123, and B135.

[52] Henrich (1989), pp. 271–2. Carl (1997) diagnoses Henrich's confusion.

The alternative, formal, interpretation involves denying that when Kant gives priority to the unity of consciousness over the unitary self, reductionism is the consequence. But the formal reading of the consciousness of self-identity – one that claims that it is necessary for experience that the subject presuppose its own identity – must be correctly elucidated in order to avoid the error of a reductionist interpretation. The formal reading must not suggest that there is a fact of the matter about the identity of the subject, independent of the alleged subject's own presupposition. This mistaken view is assumed for instance by Bermúdez, who regards the unity of apperception as the regulative principle that "it must be possible for all members of a suitably unified set of experiences to be ascribed to a single numerically identical subject" – a principle that does not entail that there really is a single numerically identical subject: "It might well be the case, as Kant hypothesises, that what seems to be a single numerically identical subject turns out to be a causally connected series of distinct subjects…"[53] His interpretation of the formal reading, which assumes that there is a fact of the matter about the identity of the subject, independent of the alleged subject's own presupposition, attempts to distinguish a putative unity from an actual one – a distinction that makes sense only on the reductionist assumption that there can be a "suitably unified set of experiences" that is not the set of experiences belonging to one individual. The "regulative principle" can only be an expression of IEM. There is no other guarantee of the persistence of the subject and the unity of consciousness, even if in some sense it must be assumed.

Nor does Kant hypothesise that a subject that appears identical may not in fact be so, as Bermúdez also claims. This is shown by the neglected anti-Cartesian context to the Third Paralogism – the Paralogism also cited by Evans in his warning about the danger of tautology in the treatment of self-consciousness.[54] Recall the Paralogisms' leitmotif: "The analysis…of the consciousness of myself in thought in general, yields nothing whatsoever towards the knowledge of myself as object. The logical exposition of thought in general has been mistaken for a

[53] Bermúdez (1994), p. 228. Allison (1983) also proposes a formalist reading, see p. 140.

[54] Another example is Brook (1994). Sacks (1991) is notable in rejecting the standard interpretation. Cassam's extreme formalist account holds that Kant regards the "logical identity of the 'I'" as "insensitive to the boundary between one thinker and another" (1997a, p. 159). This seems not to capture the idea of "change" in the "I", but in any case does not support the standard view.

metaphysical determination of the object" (B409). That is, as Strawson neatly expresses it, Rational Psychologists mistake the unity of consciousness for consciousness of a (Cartesian) unity.[55] This analysis leads Kant in the Third Paralogism to contrast the guaranteed "identity of the consciousness of myself at different times" – the "logical identity of the 'I'" – with the "numerical identity of my subject", which is not guaranteed. Both Bermúdez and Evans assume that the "logical identity of the 'I'" is a fact of the matter concerning personal identity – and that in rejecting it, Kant endorses reductionism. In fact, the guarantee of numerical identity which Kant rejects is a Cartesian notion – one of the illusions of Rational Psychology that he wishes to unmask, and implicit in the claim that "We … connect all the appearances, all the actions and receptivity of our mind, *as if* the mind were a simple substance which persists with personal identity … ".[56] In the Third Paralogism, "person", "identity", and "subject" all have a Cartesian sense: "person" is equivalent to "Cartesian thinking subject". Kant's discussion thus echoes the illustrious Locke, rather than anticipating the renowned Parfit. This is apparent when one recognises that reductionism is not the only alternative to the Cartesian unitary self.

In order to defend the non-reductionist interpretation that I am proposing, passages such as the following must be treated with care:

> The identity of the consciousness of myself at different times is therefore only a formal condition of my thoughts and their coherence, and in no way proves the numerical identity of my subject. Despite the logical identity of the "I", such a change may have occurred in it as does not allow of the retention of its identity, and yet we may ascribe to it the same-sounding "I", which in every different state, even in one involving change of the [thinking] subject, might still retain the thought of the preceding subject and so hand it over to the subsequent subject. (A363)

Kant then has a curious footnote:

> An elastic ball which impinges on another similar ball in a straight line communicates to the latter its whole motion … If, then, in analogy with such bodies, we postulate substances such that the

[55] This claim of the Paralogisms is pursued in Chapter 5.
[56] "Appendix to Transcendental Dialectic", A672/B700.

> one communicates to the other representations together with the consciousness of them, we can conceive a whole series of substances of which the first transmits its state together with its consciousness to the second, the second its own state with that of the preceding substance to the third ... The last substance would then be conscious of all the states of the previously changed substances, as being its own states ... And yet it would not have been one and the same person in all these states.

Parallels with Locke's separation of the self or person from a Cartesian thinking substance are too strong to be accidental. Locke writes: "if the same consciousness ... can be transferred from one thinking substance to another, it will be possible that two thinking substances may make but one person. For the same consciousness being preserved, whether in the same or different substances, the personal identity is preserved".[57] One should note that when Kant writes that "the last substance would then be conscious of all the states of the previously changed substances, as being its own states", he does not add "mistakenly so", as a reductionist interpretation requires. Rather he says, "And yet it would not have been one and the same person in all these states" – and since we are told that the balls are a "series of substances", "person" should be taken as "Cartesian thinking thing".

Evans is therefore wrong to suggest that it is the danger of tautology in descriptions of self-identification that Kant has in mind in the Third Paralogism, if this implies that Kant is concerned to contrast "objective" personal identity with "subjective" self-identity. The example in Kant's footnote is intended to show the irrelevance of the Cartesian thinking substance to an understanding of self-consciousness – not the possible delusion of the subject in postulating its own persistence. The same interpretation applies to the main text. The "change [that] may have occurred in [the 'I'] as does not allow of the retention of its identity" is a change of Cartesian thinking substance, not of personal identity in the ordinary sense. This reading disposes of the most significant apparent textual support for the view that when Kant gives priority to the unity of consciousness over the unitary self, reductionism is the consequence.

The same confusion of Kantian and reductionist conceptions of the unity of consciousness is found in J. S. Mill's presentation and subsequent retraction of a phenomenalist construction of the self. Although

---

[57] Locke (1975), Book II, Chapter 27, Section 13.

it is unclear why he withdraws the series account, he cites the reason that the phenomenon of memory "[cannot be] adequately expressed, without saying that the belief [it includes] is, that *I myself* formerly had ... the sensations remembered" (my italics). Later, he claims that the series account of the self neglects the "real tie ... which connects the present consciousness with the past one, of which it reminds me ... [as opposed to] a mere product of the laws of thought without any fact corresponding to it".[58] It may appear that Mill is referring to the phenomenal unity of the stream of consciousness, but in fact he is alluding to the necessary unity of the remembering or anticipating subject, which concerns experience reminisced and the experience of reminiscence. Mill is looking for some principle of the necessary unity of consciousness. The "real tie" would allow that the subject's feeling of identity tracks a pre-existing fact of the matter – hence Mill renounces the unity of consciousness standpoint in its Humean sense for that of the unitary self. However, it is IEM which Mill is groping for in his postulation of a "real tie", while in his other writings, as we have seen, he anticipates the claim of a conceptual holism between memory and personal identity.

The discussion of these issues is concluded in Chapter 5, which concerns proprioception and bodily identity. That chapter undermines the frequent objection to the unity of consciousness standpoint: that it neglects the subject's embodiment. The complaint may be true of Hume's, and to a lesser extent Kant's, discussion, though it applies less to phenomenalists such as Mill, James, and Ayer, who proposed to construct the self and the external world – in which they included the body – in parallel. It does not apply to the account offered in the present volume.

APPENDIX: De Gaynesford's critique of IEM as a merely pragmatic phenomenon

A recent deflationary treatment of IEM by Max de Gaynesford rejects the explanation of what he calls the "nonsense-question phenomenon", that is, IEM, in terms of "I"'s identification-independence (what he calls *Independence*). His initial suspicion of the lack of consensus on what *Independence* is, indicating the instability of the doctrine, can be dismissed; such a situation is common in areas as difficult as self-

[58] Mill (1979), pp. 194, 207. In fact, the reasons for Mill's retraction of the series account are not at all clear; the question is discussed in Hamilton (1998a).

consciousness and self-reference.[59] De Gaynesford also assumes that proponents of IEM regard "subject" uses of "I" as central. That is not a claim of this volume, as we see in the final chapter, which holds that discrimination and non-discrimination requirements involve the interdependence of "object" and "subject" uses.

A more compelling objection is that IEM is insubstantial and explicable pragmatically by reference to the convention that one asks "whether one has got the reference right only when there is ... an evident possibility or likelihood of one's having got the reference wrong".[60] This convention rests on the more general Gricean convention that speakers must be relevant and perspicuous in their discourse. De Gaynesford argues that because the question is nonsensical only on certain occasions – what I have termed the relativity of IEM to a ground – the nut to be cracked does not require as massive a hammer as *Independence*. He then cites counter-examples to IEM – such as "deviant causal chains" giving rise to q-proprioception – which were diagnosed in Chapter 2 as question-begging. He concludes that the "unimposing" phenomenon of IEM requires an "equally trivial" pragmatic explanation: "we have good reason to avoid redundancy in reporting on our own states to ourselves and to others ... [this] means not saying who or what has been referred to when it is perfectly obvious".

The most important line of argument here is the familiar anti-Wittgensteinian position that, as de Gaynesford puts it, "If there are situations – no matter how atypical – on which it is sensible to ask these questions and demanding an answer to them, then it would *never* be strictly nonsensical to ask them. On many or even most occasions, these questions might be ignored or answered without difficulty".[61] That fact, de Gaynesford holds, can be explained pragmatically. The claim is that if something makes sense in some circumstances, it makes sense in all, though on occasion it might be unreasonable or even irrational to *say* it; for instance, since "I doubt that I have two hands", said by a patient in hospital with severe frostbite, makes sense, then the utterance always makes sense.

As so often in discussions of Wittgenstein, a mere statement of the contrary position is taken as a refutation of his view. In defending his position, one must obviously examine individual cases with care; many expressions which serve no useful purpose on a particular occasion

[59] De Gaynesford (2006), pp. 51–5.
[60] De Gaynesford (2006), p. 59.
[61] De Gaynesford (2006), p. 62.

undoubtedly do make sense. At *On Certainty* paragraph 10 we see Wittgenstein wrestling with these issues:

> I know that a sick man is lying here? Nonsense! I am sitting at his bedside, I am looking attentively into his face. – So I don't know, then, that there is a sick man lying here? Neither the question nor the assertion makes sense… And "I know that there's a sick man lying here", used in an unsuitable situation, seems not to be nonsense but rather seems matter-of-course, only because one can fairly easily imagine a situation to fit it, and one thinks that the words "I know that…" are always in place where there is no doubt, and hence even where the expression of doubt would be unintelligible.

This line of argument requires a careful defence that I develop elsewhere.[62]

De Gaynesford also has an interesting line of objection to the Wittgensteinian view that IEM judgements do not involve genuine identification. He argues that *Independence* threatens the requirement that criteria of personal identity underlie a subject's self-ascription of current or personally remembered states of consciousness, and thus severs the links between uses of "I" and empirical criteria of personal identity. This question is raised by Strawson in his discussion of an early version of IEM, where he argues that "I" can be used without criteria of identity and yet refer to a subject because it issues from the mouth of a person identifiable by the application of empirical criteria of personal identity, and is "used by a person who would acknowledge the applicability of criteria [of personal identity] in settling questions as to whether he, the very man who now ascribes to himself this experience, was or was not the person who, say, performed such-and-such an action in the past".[63]

De Gaynesford responds that the use will not self-refer if the person is being used merely as a mouthpiece; and that Strawson confuses the semantic issue of whether "I" may be used to express thoughts independently of identification and still count as a referring term, with the pragmatic issue of what an audience requires in order to understand the speaker's utterance. Furthermore, de Gaynesford argues, by assuming that the "I"-user is a person, Strawson begs the question against the possibility of a sub-personal "I"-user. This line of objection, as well as

---

[62] Hamilton (forthcoming 2013), Chapter 3.
[63] Strawson (1967), p. 165; De Gaynesford (2006), pp. 66–7.

returning to the pragmatic-semantic contrast just discussed, also raises the question of non-personal and thus non-referring uses of "I". Such uses are parasitic on the core, personal use of "I" as a device of self-reference, I believe, and so cannot supply the basis for substantive arguments against IEM.

4

Proprioception and Self-Consciousness (1): Proprioception as Direct, Immediate Knowledge of the Body

We now turn from memory to proprioception. This chapter and the following one apply the same treatment to proprioception and bodily identity, involving conceptual holism, as that applied to memory and personal identity. The faculty or capacity of *proprioception* is both familiar – because it underlies the possibility of action – yet mysterious. It yields ordinary knowledge of bodily position and movement – what is loosely termed "bodily awareness". Yet in Philosophy it has until quite recently been neglected; indeed, in my experience the issues it raises remain unfamiliar to general philosophical audiences. Hence a rather fuller account of the nature of proprioception is required than in the case of memory. This chapter aims to demystify proprioception by considering both Phenomenological and Gibsonian accounts. There is important common ground between Gibson's position and that of Phenomenology, both influenced by Gestalt psychology, and a philosophical treatment of the body and self-consciousness should draw on each.

The varieties of proprioception are complex. The core capacity yields knowledge of bodily position and movement. Strictly speaking, *kinaesthesis*, often used as equivalent to proprioception, is knowledge of movement of parts of the body, as opposed to their posture or position.[1] Other varieties of proprioception include knowledge of fatigue and warmth and cold (as opposed to merely feeling tired, hot, or cold); the

[1] The confusion is perhaps understandable, as knowledge of movement requires knowledge of position. Cole (1991), pp. xix–xx, corrects the error.

inner ear's vestibular system that gives information about balance and posture; interoception (the visceral sense); and "visual proprioception", the term coined by J. J. Gibson for the kinaesthetic function of vision, enabling the subject to differentiate between a change of place by the observer, reversible by moving back to the original position of observation, and a change of state of an external object.[2] Interoception yields knowledge of the non-muscular organs, blood-vessels, and intestines, and is a function of the autonomic nervous system.[3] Proprioception also includes self-locating capacities that yield more than mere internal or surface bodily knowledge – such as knowledge of orientation in space, for instance that I am lying down or standing up.

Considering cases of proprioceptive deficit assists the imaginative effort of trying to grasp what proprioception involves. What is often termed *proprio-blindness* – though we later see why the description is tendentious – is caused by conditions ranging from stroke to viral infection, and is more common, at least in a partial and mixed form, than is often imagined.[4] Everyone has experienced the very limited proprioceptive deficit of their arm going numb after sleeping on it. Ignorance of the prevalence of proprioceptive deficit is partly the result of failure to recognise the distinct neuropathies of the efferent and the afferent systems; patients with proprioceptive deficit suffer from the latter but not always the former. Stroke victims usually suffer from a complex mixture of damage to both efferent and afferent systems; some parts of a limb may be truly paralysed, while other parts exhibit loss of proprioception, temperature awareness, or sensation of pain. Patients with proprioceptive deficit may be only effectively paralysed because, as neurologists put it, the brain cannot tell a muscle what to do if it does not know where it is.

Devastating cases of proprioceptive deficit such as Ian Waterman's are rare.[5] At the age of 19, an immune reaction ravaged his nervous system

---

[2] Gibson (1966), pp. 37–8. A more complete list of the varieties of proprioception is found in Bermúdez et al. (eds) (1995), p. 13, and Bermúdez (1998), pp. 132–3.

[3] Some researchers maintain that one can learn to distinguish visceral events just as well as muscular ones.

[4] The experiment is discussed in Hamilton (1991) and below, Chapter 6.

[5] His heroic story is told in Cole (1991), and in a BBC2 TV programme, broadcast on 16 October 1997. Cole writes that "damage to his nerves was extraordinarily, perhaps uniquely, specific. It had affected some of the sensory fibres, but none of the motor nerves" (p. 2; further physiological details are found on pp. 24–34). Sacks (1985) has an earlier discussion of a patient with severe proprioceptive deficit. I am indebted to Ciarán Benson for discussion of his experience of proprioceptive deficit.

from the neck down, while leaving the vestibular system and visual proprioception still functioning. Waterman was left a helpless "rag doll", who had to be fed, washed, and dressed; attempts at movement elicited only uncontrolled jerks. Unlike most other patients with severe proprioceptive deficit, however, he re-acquired mobility through the constant effort or the "daily marathon" of a mostly visual tracking of his body. His knowledge of bodily position and posture is sustained by constant looking and checking. He retains a capacity to feel pain, temperature, and muscle fatigue, and presumably has some knowledge of the location of pain, at least through primitive reactions of pain response such as movement. Without this residual proprioception his extraordinary achievement in regaining bodily control and learning to walk again would have been impossible. But even after 30 years of intense practice, the simplest movement has not been automated, but requires concentrated, strenuous visual attention. In the dark he still collapses like a rag doll.

The plight of Ian Waterman gives some insight into the contrast between normal bodily awareness and knowledge of bodily position and movement based on visual perception, a contrast fundamental to a central argument of this chapter and the next:

> *IEM of bodily self-ascription generates a challenge to materialist accounts of bodily identity, showing that these should be rejected in favour of a self-conscious knowledge account.*

Judgements such as "My legs are crossed", made on the basis of bodily awareness, exhibit IEM. Thus there are unexpected parallels between proprioception and memory, not least in how materialist and imagist accounts of it have prevented a proper understanding of IEM.

IEM underlies a *self-conscious knowledge account* of bodily identity, which says that "my body" is the body of which – when I am conscious – I have self-conscious knowledge, and which I can move in a basic sense. On this account, X is a body part of mine if and only if I have IEM-exhibiting knowledge of it, and can move it basically. (This formulation is qualified later, to accommodate body parts of which we do not have proprioceptive awareness.) In contrast, *materialism about bodily identity* holds that a body part is mine if and only if it is part of the material unity that I am.

The present chapter advances this core argument through the following considerations:

i) Phenomenological and Gibsonian accounts are presented as alternatives to materialism about the body, the default position of modern

philosophy, and to the associated perceptual model, according to which proprioception is a mode of perception that tracks a material entity.

ii) Both the image theory, and the perceptual model, are rejected. The *image theory* treats proprioceptive knowledge as inferential; the ubiquitous *sensory or perceptual model* assimilates proprioception to the five senses, treating it as a "sixth sense" or "muscle sense" of bodily awareness. This chapter argues, in contrast, that proprioception, like memory, yields a kind of direct, non-inferential knowledge. One does not have to do anything to acquire it; I "just know" that my legs are crossed. Proprioception constitutes *basic knowledge* of one's body in the same sense that *one moves one's body basically* – as in cases of moving my arm, where I do not move it by doing something else, such as lifting it with my other arm. One could say that proprioception is a *non-sensory but primordial capacity*.

iii) IEM is shown to be exhibited by bodily as well as psychological self-ascription, in the same substantial way exhibited by memory. Distinctively self-conscious knowledge – that which necessarily yields knowledge only of the subject – is manifested in IEM judgements of posture, orientation, intention, and action. This defence of IEM is developed in the following chapter into a critique of materialism about bodily identity and ownership, and a self-conscious knowledge account is presented in its place.

4.1 The Phenomenological challenge to materialism about the body

This and the following chapter target materialism about the body, and the related position of materialism concerning bodily identity, presenting a critique that originates in the Phenomenological tradition. What I term *materialism about the body* is an almost universal post-Cartesian assumption, which says that the body is a purely material and epistemically outer entity, whose states, unlike those of the mind, are not known immediately and transparently. It is not just materialists in the ordinary sense who are committed to this position. Almost all philosophers, dualist and materialist, subscribe to it, except those from the Phenomenological tradition of Husserl and Merleau-Ponty, and their Analytic affiliates.[6] Locke and Schopenhauer are interesting exceptions,

[6] Idealists begin with the assumption that the body is material, and then try to construct or construe it in ideal terms.

as we see in Chapter 5. Bodily awareness is the Archimedean crux of Schopenhauer's philosophy, though he mistook the non-phenomenal or non-perceptual nature of proprioception for an intimation of knowledge of the thing-in-itself.[7]

Despite interesting precedents in Locke as well as Schopenhauer, Husserl was the first to describe a physical-intentional ambiguity concerning the body, which suggests that the sensory model of proprioception may not be correct. His account is a great advance on that of his precursors. In *Ideas Book II* he argues that the human body is defined by intentional attributes of action and proprioception as well as by spatio-temporal material attributes:

> ... the Body is originally constituted in a double way: first, it is a physical thing, *matter*; it has extension, in which are included its real properties, its color, smoothness, hardness, warmth ... Secondly, I find on it, and I *sense* "on" it and "in" it: warmth on the back of the hand, coldness in the feet, sensations of touch in the fingertips ...
>
> [The Body is] a material thing which, as localisation field for sensations and for stirrings of feelings, as complex of sense organs, and as phenomenal partner and counter-part of all perceptions of things ... makes up a fundamental component of the real givenness of the soul and the Ego.[8]

Husserl distinguishes two senses of "the body", *der Leib*, the "animated flesh of an animal or human being" – the "lived Body" or what may be referred to as "my body" – and *der Körper*, "inanimate physical matter", the "mere" body or body viewed purely as a physical object.[9] This contrast is central to the discussion in this and the next chapter, which emphasises the priority of the lived Body, neglected before Husserl.

The lived Body is sometimes referred to as the "mindful body"; a better alternative might be "the body-in-action".[10] These terms are useful chiefly as indicating essential features of, simply, my body. Although Husserl's arguments help to undermine materialism about the body, recent commentators, portraying him as a thoroughgoing

---

[7] See Janaway (1999), Chapter 8.
[8] Husserl (1989), Sections 36 and 40, pp. 153, 165.
[9] Husserl (1989), for instance Section 62, p. 297.
[10] It has also been termed the "living Body" by Cassam (1997a), p. 52, and by Bell (1990), p. 208, which has misleading biological connotations. MacDonald addresses these issues in his (1996) and (1999).

anti-Cartesian, overstate his rejection of Cartesianism. The metaphor of animation of my body, and reference to the "real givenness of the soul and the Ego", shows a residual Cartesian bias. Cartesianism, in the textbook sense in which I am using the term, involves both materialism about the body and a version of the *control model of action* in which the body is directed by an immaterial ego. Even in Merleau-Ponty's presentation of the "body-subject", the Cartesian model of action as bodily control continues to exert an influence, as we will see. In *The Structure of Behaviour*, Merleau-Ponty expressed dissatisfaction with the residue of dualism in his earlier work, in particular in his treatment of the "bodily ego".[11]

The ideas of these Phenomenologists, filtered through the work of Gareth Evans, have achieved some currency among Analytic philosophers, except among supporters of scientistic naturalism.[12] But the implications for materialism about the body have not been assimilated. Even Evans, as we saw, has affinities with materialism. He recognises that the mental-physical distinction is, as he puts it, "almost entirely arbitrary", which is absolutely right; the simple dichotomy of mental and physical is too crude, as the ambiguous status of the body crucially illustrates. However, Evans also claims that persons are fundamentally identified as material entities.[13] Merleau-Ponty's Phenomenological project, in contrast, posits stricter limits for scientific psychology, and his account of bodily intentionality is more radical than Evans's in its

[11] Merleau-Ponty (1962); see also B. Smith (1995), in B. Smith and D. Smith (eds. 1995), pp. 406–7. D. Smith (1995, pp. 324–6) argues that the break with Cartesian dualism in Heidegger and Merleau-Ponty develops Husserlian insights. Husserl may also have an animalist bias: the biologistic stance that treats persons as essentially human animals. *Der Leib* and *der Körper* parallel Locke's distinction between man and person; the "Lived Body" may be a whole made up of two dependent wholes, mind and *Körper*.

[12] Evans (1982). Husserl and Merleau-Ponty are unmentioned in Bermúdez (1998) and Baker (2000). Merleau-Ponty influenced Evans via Charles Taylor; Evans (1982), p. 156, quotes a passage where Taylor discusses Merleau-Ponty. Other Analytic writers influenced by Husserl and Merleau-Ponty are Dreyfus and Dreyfus (1999); Dreyfus (2000); Kelly (2002).

[13] Evans refers to "an almost entirely arbitrary, though traditional, distinction between the mental and the physical" (ibid. pp. 222–3). "I am not that hard up for categories", Wittgenstein retorted when asked whether there is one substance or two, rejecting substance ontology, not affirming metaphysical pluralism (Wittgenstein 1980, vol. ii, para. 690). There are clearly mental entities (thoughts) and clearly physical ones (stones, electrons), but others are not clearly either (emotions, the lived Body).

claim that the experience of one's own body "reveals … an ambiguous mode of existing".[14] The considerations raised by Husserl and Merleau-Ponty, I believe, support the non-materialist, self-conscious knowledge standpoint that "My body" is the body of which, when I am conscious, I have self-conscious knowledge, and which I can move basically.

Associated with materialism about the body is the sensory model of proprioception dominant among Analytic writers, which treats proprioception as a mode of perception that tracks a material entity. The perceptual model assumed by Sherrington remains almost ubiquitous in physiology and psychology. In *The Integrative Action of the Nervous System* (1906), Sherrington coined the term proprioception – as a contraction of "proprioperception", that is, "self-perception" – and called it our "secret sixth sense". He distinguished three main groups of sense organs: exteroceptive, such as those that detect light, sound, odour, and tactile stimuli; interoceptive, exemplified by taste receptors; and proprioceptive, or those receptors that detect events occurring in the interior of the organism.[15]

This chapter is concerned to develop the alternative model suggested by Phenomenologists. A further source of a non-sensory approach is the work of psychologist J. J. Gibson. Gibson formulated his view of proprioception in reaction to that of Sherrington and his contemporaries, who assumed that "each sense had to have its specialised receptors that could excite corresponding sensory nerves" – a version of the perceptual model.[16] In the 1960s, Gibson argued against the traditional idea of proprioception as a specific "body sense", referring instead to self-specifying information which cuts across the different sensory modalities.[17] He treats proprioception as "a component of the functioning of all the perceptual systems", and claims that proprioception and perception are interdependent – the *interdependence thesis*.[18] One of his important

[14] As Merleau-Ponty writes in the "Preface" to *Phenomenology of Perception*, "Husserl's first directive to phenomenology, in its early stages, to be a 'descriptive psychology', or to return to the 'things themselves', is from the start a forswearing of science. I am not the outcome or the meeting-point of numerous causal agencies which determine my bodily or psychological make-up" (Merleau-Ponty 1962, p. viii).

[15] Sherrington (1973).

[16] Gibson (1968), p. 33.

[17] Gibson (1966, 1968), on which see Reed (1988). Husserl had earlier commented on the interwovenness of perception and bodily movements, discussed by B. Smith (1995), p. 403.

[18] Reed (1988), p. 227.

insights was that visual perception contains proprioceptive information essentially, as well as exteroceptive information about the environment.[19] I return to Gibson's interdependence thesis later, in order to clarify its relation to the non-perceptual account, seeking to undermine the view that proprioception is a capacity for tracking an essentially material entity.

4.2 Proprioception as direct, non-inferential knowledge: rejecting the image theory

It is natural to think of proprioception in sensory terms. But here I will criticise the sensory model, arguing that proprioception is definable by the direct character of the knowledge that it yields. It yields direct, immediate, and spontaneous knowledge of the body – centrally its position, posture, and movement. This knowledge is immediate not just in the sense that it is non-inferential, but also in that the subject does not have to do anything to acquire it. Indeed, I do not really "acquire" it at all; I "just know", for instance, that my legs are crossed when they are. So proprioception differs both from knowledge based on bodily sensation, and from perception of the body by the five senses. Proprioceptive knowledge such as awareness of fatigue and temperature must therefore be distinguished from avowals of bodily sensation. However, the focus in the present discussion will be on the central cases of proprioceptive knowledge, namely, position, posture, and movement.

The characterisation of proprioceptive knowledge as direct, immediate, and spontaneous is preferable to Anscombe's term "non-observational", which appears in her influential discussion of bodily awareness. Non-observational knowledge, she writes, includes knowledge of one's own intentional actions and of the causes of some involuntary movements, as well as knowledge of the position of one's limbs – that is, proprioceptive knowledge. This knowledge is non-observational, she argues, because

> nothing shows [someone] the position of his limbs; it is not as if he were going by a tingle in his knee ... Where we can speak of separately describable sensations, having which is in some sense our criterion for saying something, then we can speak of observing that thing. ...[20]

[19] Gibson (1979).

[20] Anscombe (1963), p. 13. Like other earlier writers, she refers to "bodily awareness" rather than "proprioception".

Anscombe is right to reject a separately describable basis for proprioceptive judgement – an issue we return to later – but her terminology is unsatisfactory. While "non-observational" implies a contrast with perceptual knowledge, her elucidation of the term seems to contrast it with inferential knowledge based on bodily sensation.

Anscombe is wrong to say that, in itself, observation implies "separately describable sensations", since this suggests that it is inferential. If this were so, knowing that my legs are crossed by looking could also count as non-observational, since it does not seem to involve "separately describable sensations". This makes nonsense of her criterion. The error results in a conflation of two distinctions: between inferential and non-inferential knowledge, and between (self-) knowledge by the five senses and distinctively self-conscious self-knowledge (knowledge which necessarily concerns only the subject). Both contrasts must be made explicit and each, I will argue, involves a distinct challenge to the direct knowledge account, from the image theory and perceptual model respectively. There are further problems with Anscombe's terminology, for it is not even the case that "non-observational" implies "non-perceptual". Despite philosophy of science's concern to demarcate a class of observation sentences, there is something artificial about equating ordinary perception with observation, since "observation" implies attentive perception. The characterisation of proprioceptive knowledge as "direct, immediate, and spontaneous" is therefore greatly preferable to Anscombe's term "non-observational".

My claim is that proprioception is not inferential knowledge, and that neither should it be assimilated to knowledge by the five senses. Rather, it is a *non-sensory but primordial capacity*. Hence I reject both the image theory and the perceptual model. The *image theory of proprioception*, analogous to the image theory of memory, maintains that proprioceptive knowledge is inferential. It is a representational theory, and assumes that proprioception may be distinguished phenomenologically, by its feel – by the sensations associated with it; the presence of these sensations is the inferential basis of the proprioceptive judgement of bodily posture or position. Proponents of this view take the concept of a "body-image" – the subject's total proprioceptive knowledge – literally. The term is more appropriately applied to the highly distorted self-image possessed by anorexics and some psychotic patients.[21] It is important for the treatment of self-consciousness that the image theory should be

[21] Cole and Paillard (1995) in Bermúdez et al. (eds) (1995) has further discussion of "body-image".

discredited, since – like the perceptual model – it makes q-proprioception appear more plausible, just as the image theory of memory makes q-memory more plausible. Q-proprioception is meant to be a peculiar kind of proprioception, involving an information-link between the subject and someone else's body, such that the subject allegedly registers information from it, thus contravening the IEM of proprioception. (The phenomenon is defined more closely at the end of this chapter.) The image theory makes it seem more imaginable that the proprioceptive "feel" could arise from the bodily state of a distinct subject.

The picture offered by the image theory is this: if my legs are crossed with one resting on the other, I experience feelings of pressure, touch, and so on; even when my arm is by my side but not touching anything, there will be feelings of muscle tension or skin stretching, or tension in my shoulder. Such sensations "advise" me of the positions of my limbs, and so forth, presumably through inductive inference. For this account to be plausible, the sensations would have to be characterised independently of the state of whose occurrence they advise. This is possible when I infer from bodily sensations to the existence of a medical condition, having learned from experience that the feeling is associated with this condition, for instance when I infer, from the recurrence of stomach pains, that I have a gastric ulcer rather then indigestion. But as Anscombe and Wittgenstein argue, in the proprioceptive case it is difficult to see how sensations could be characterised independently of the state of whose occurrence they advise. Wittgenstein's discussion suggests an interpretation of the "feeling" involved in proprioception which undermines both the perceptual model and the image theory; we will return to this later. Here I address considerations that seem to support the image theory.

The first is the suggestion that acquiring a learned capacity has to involve inference, and that information-processing amounts to inference. Clearly, children learn to control their movements, but an inferential model of this learning process, in particular the suggestion of unconscious inference, is unconvincing.[22] In adult learning processes, conscious inference is more plausible. In Alexander Technique, for instance, one learns to attend closely to one's bodily posture, aiming to release bodily tensions and direct bodily use efficiently, by "thinking" upwards when sitting down, for instance. The habits that are encouraged

[22] See Budd (1989), pp. 147–9. Other criticisms of an image theory are found in Anscombe, "On Sensations of Position" in her (1981a), and Candlish (1996).

at first feel strange, but gradually become automatic. For instance, I may learn on the basis of close attention that I have a detrimental habit of stooping forward, even though I feel that am standing quite upright. Alexander Technique attempts to make this knowledge immediate and habitual, by means of a process that does not seem to involve conscious inference; the very elusiveness of the technique counts against the idea that inductive inference is involved.

In other cases knowledge of bodily posture and movement is not immediate and has to be learned. In learning any new skill, sport, or art, one has to integrate proprioceptive input, and the proprioceptive capacity can be sharpened through study of different disciplines. Visual perception or touch is often essential in refining a rough and approximate proprioceptive knowledge. Figure skating, ballet, playing a musical instrument, and singing are examples. For instance, proprioception is not always accurate enough for such specialised purposes as learning complicated dance positions, and one has to use mirrors. If I am instructed to close my eyes and put my arms out horizontally, I will manage it near enough; knowledge that my arm is horizontal, based on proprioception, means "horizontal as opposed to vertical or at 45 degrees". But singing or speaking a foreign language with a good accent require more precise bodily knowledge, such as knowing where my tongue is when making certain sounds. Before the knowledge becomes automatic, the learning process involves exploratory touch: "My tongue is touching my upper teeth" is more like "My fingers are in the jelly" than "My fingers are curled up". (Exploratory touch occurs when I move my hand to feel part of my body or another object; passive or proprioceptive touch occurs when I feel the rain on my face, on which more below.) Ian Waterman learned to control his body by a process of constantly looking, which never became automatic or unconscious, and must often involve inferences such as "When my body is at this angle, it's likely that I'm about to topple over".

4.3 Proprioception as direct, immediate knowledge: rejecting the perceptual model

The image theory and the perceptual model have many proponents, and it is a difficult question how they are related.[23] As noted earlier, proprioception has been called a "sixth sense" or the "muscle sense",

[23] In their criticism of Anscombe, Bermúdez et al. (eds) (1995), pp. 18–19, run the two together.

and physiological affinities with touch seem to support the perceptual view. The proprioceptive nerve receptors in the muscles might be regarded as "organs of proprioception" and "organs of touch": they give feedback from joints, tendons, and muscle spindles, while cutaneous receptors near the skin surface respond to touch. If there is an organ of exploratory touch, one of the five senses, then it seems that there must be an organ of proprioception. However, I will argue that the divergence in the character of knowledge yielded by proprioception, and by the five senses, undermines the standard conception of proprioception as perceptual knowledge, contingently limited in its objects to the body and the bodily surface. Rather, one should regard it as a non-sensory but primordial capacity.

It may be felt that nothing decisive depends on whether proprioception is regarded as a mode of perception. "Perception", like "self-consciousness", may be a philosopher's term of art; appeals to ordinary usage therefore have a restricted role in deciding whether a faculty counts as perceptual. However, the perceptual model is not innocuous. Its claim that I simply shift my attention from objects outside one of my bodily boundaries to what is going on at or beneath that boundary, makes it seem contingent that I have proprioceptive knowledge only of my own body. This claim fails properly to acknowledge proprioceptive IEM. Although proponents of the perceptual model often acknowledge that proprioception is unique among modes of perception in providing knowledge only of one object and its parts, they cannot explain why this is. Bermúdez attempts such an explanation, arguing that the perceptual model meets the constraints that perception must involve identification or tracking of objects over time, and that ordinary modes of perception involve the perceiving of a multiplicity of objects. However, satisfaction of the latter constraint, while allowing Bermúdez to reject the possibility of q-proprioception, implausibly broadens the extension of self-consciousness.[24] The perceptual model denies the conceptual interdependence of proprioception and bodily identity and thus embodies the residue of a traditionally Cartesian picture of bodily awareness. Despite the disclaimers of some proponents, therefore, it does underlie the science-fictional scenario of q-proprioception. In contrast, the direct

[24] Bermúdez (1998), pp. 137–51, especially p. 144. Another example is Martin (1995), who argues that an account of "what ties the content of bodily awareness to a particular object … does not seem to be one that we can provide purely a priori" (p. 283).

knowledge account is fully consistent with the concept of a body-subject. (These issues are pursued in the next chapter.)

In assessing the perceptual model, two questions arise: how does proprioception differ from the five senses, and does this difference show that it is not a variety of perception? My conclusion is that proprioception has a status intermediate between, on the one hand, the five senses which yield perceptual knowledge, and, on the other, awareness of sensations such as pain which give rise to avowals, and which are not – I would argue – objects of knowledge at all. Thus the epistemic status of "I am cold" is intermediate between "the ice is cold", known on the basis of touch, and "I feel cold" (or "I have a headache") – though it is closer in certain respects to sensation. Some writers assimilate proprioception to sensation by subsuming both under the perceptual model, thus neglecting the well-grounded distinction between sensation and observation. Their position re-states the Cartesian model of sensations, which treated them as objects of inner perception, implausibly postulating a sensational object distinct from the awareness. The contemporary version regards pain not as an object of perception, but as a mode of perceiving one's body, thus pain has a physical object distinct from the awareness. Martin for instance claims that the object of pain is the body part that feels painful, just as the object of visual experience is the physical object that one is perceiving: "[O]ne perceives one's body through sensation, just as one perceives other objects through the five senses".[25]

I wish to reject these treatments. One must surely take seriously the traditional distinction between sensation and perception. The most that can be conceded to those who do not is that the self-ascription of pain assumes proprioceptive knowledge, and that, as noted earlier, experience of pain may itself be the basis for inferential knowledge, for instance that I have a gastric ulcer. Aside from its innate implausibility, a perceptual model cannot account for the immunity to error exhibited by avowals of pain. Their authority implies that if X truthfully, attentively, and comprehendingly asserts or avows "I have a pain in my leg", and they have no relevant cognitive defect, then the truth of "X has a pain in their leg" is guaranteed.[26] I can be mistaken about the causal origin of the pain, but not about its phenomenal location; it is the latter which conditions the truth of the avowal. It follows that apparently

[25] Martin (1995), p. 269.

[26] The implications of the authority of avowals are developed in Hamilton (2000).

mistaken self-ascriptions resulting from referred pain and phantom-limb phenomena, often cited as counter-examples to the authority of avowals, are not convincing. An example of *referred pain* would be a sinus condition which causes a headache which is not phenomenally located in the sinus region, but the fact that the headache has its immediate causal origin elsewhere does not make the avowal that I have a headache a mistaken one. A similar point applies against objections that cite *phantom-limb phenomena*, where the subject whose left leg has been amputated is inclined to avow "I have a pain in my left leg". Certainly they have a pain, and it is phenomenally located where the left leg would have been.

Although the distinction between sensation and observation is well-grounded, proprioception seems to belong to neither category. If treated simply as a mode of knowledge of one's body, the perceptual model seems inevitable. Nonetheless, it is possible to make a principled distinction between proprioception and sensory perception. An initial suggestion is that sensory knowledge involves reference to a possible action of looking, listening, tasting, smelling, or touching, while no kind of action is required to gain proprioceptive knowledge. Only exceptionally, where the subject has to look or touch to find out – when they have partial proprioceptive deficit due to a stroke for instance – must they do something to acquire the knowledge that their legs are crossed. (My knee may also feel as if it has a bump on it – the skin feels stretched and so on.) In the normal proprioceptive case, it is not something that one gains or acquires. I "just know" that my legs are crossed. One reason why no action is needed might be that sense organs exhibit directionality: they pick up information from a certain direction and may need to be re-oriented. I may need to focus my gaze, or turn my head, to catch what someone is whispering. Proprioception has no such directionality, no equivalent to the moving in or focusing found in exploratory touch and the other senses. *Immediate knowledge* is knowledge which I do not have to do anything in order to acquire; I "just know". For this reason, the unusual case of learning that one's legs are crossed by looking at them, although it does not involve experiencing a sensation or applying a criterion – except in the sense of ownership perhaps – does not yield immediate knowledge.

To understand the contrast between proprioception and sensory knowledge we need to develop the distinction between *exploratory* and *proprioceptive touch*. Exploratory touch is when I move my hand to feel part of my body or another object; passive or proprioceptive touch is when I feel the rain on my face, or someone treading on my toes, and

is comparable to feeling bodily sensations such as pain. Exploratory touch and visual perception, unlike proprioception, are ways in which I can also gain knowledge of someone else's bodily position, and so the first-person judgements to which they give rise do not exhibit IEM. For instance, "My legs are crossed", when known on the basis of visual perception, is not IEM. Exploratory touch, when stationary, proves hard to distinguish from proprioceptive touch, as in Merleau-Ponty's "double sensations" cited earlier: "When I press my two hands together, it is not a matter of two sensations felt together as one perceives two objects placed side by side, but of an ambiguous set-up in which both hands can alternate the roles of 'touching' and 'being touched'".[27] These double sensations simultaneously yield knowledge of one's body and one's environment. When I feel a stationary sphere in the palm of my hand, without moving my hand across its surface, both proprioceptive touch and exploratory touch – in this case stationary – seem to be involved. By means of proprioceptive touch I feel the pressure, the coldness, and the smoothness or roughness of the sphere. By means of exploratory touch I feel the shape, the hardness, and the roughness of the sphere, but I do not feel these when my hand is quite stationary. (Feeling the shape of the box is not like feeling the edge of the box, because it involves inference.) Thus exploratory touch seems to be essentially active.[28]

Does proprioception's lack of action and directionality suggest that it is not a mode of perception? A first response by proponents of the perceptual model might be that gaining knowledge on the basis of proprioception often does involve action of some sort. Perhaps I need to move my limbs about, to activate or alert my proprioceptive capacity and get coordinated. This however is not a correlate of the particular action often required with the five senses, I would argue, it is less analogous to looking and listening than to blowing my nose when I have a cold, in order to smell the food better, or shaking my head and blinking when I see something surprising. It may also be urged that, as in the case of pain, there is an action of paying attention. But while one may focus one's attention on particular sensations, this activity is a charade

---

[27] Merleau-Ponty (1962), p. 93.

[28] Ratcliffe (2008) offers a critique of the view I am suggesting, arguing that touch does not involve two distinct feelings, a feeling of the body and a feeling of something external to the body; rather, he claims, tactile experience does not always respect a clear boundary between body and world. Other helpful discussions of touch are found in Gibson (1962) and Richardson (2009), Chapter 4, and (2013, forthcoming).

if meant to gain knowledge of one's bodily position. I can often be said to know that my legs are crossed, just in the sense that I do not try to get up without uncrossing them, and thus falling over. This knowledge is immediate. Alexander Technique requires concentrating attention on one's posture, but, as noted earlier, the process is an elusive one, and it does not seem to involve an analogue of concentrating one's gaze, or straining to catch what someone is whispering.

Proponents of the perceptual model may, alternatively, concede that no action is involved in the case of proprioception, but deny the fact any significance, arguing for instance that one should not expect directionality. Compare exploratory touch. If, as proponents of a perceptual model might maintain, there is a sense organ for touch, then directionality would result from the sense organ moving to remain in physical contact with the object: thus I move my hand so that the cutaneous nerve receptors remain in contact with the object being tracked. With proprioceptive touch, the response continues, the organs are already in the place which the information concerns, and so do not need to be moved. In response I would argue that "the place which the information concerns" is within the body-subject; as shown below, the difference is vital.

The most persuasive line of objection finds exact parallels in the behaviour of perception and proprioception. If I can often be said to know that my legs are crossed, in that I do not try to get up without uncrossing them, then similarly I might know by hearing it that there is a car behind me when I cycle along the road, in that I do not try to turn right without indicating. In neither case, it may be argued, is attention required. On the other hand, the objection continues, forming the judgement "There is a car behind me" does require attention, and similarly for the judgement "My legs are crossed". There is a related objection to the distinction between exploratory and proprioceptive touch. Proprioceptive touch occurs when I feel the rain on my face, but surely, it may be argued, exploratory touch "occurs" at the same time, since I also find out about something in the world, namely, that it is raining. Why distinguish exploratory and passive touch, when for the other senses, active and passive do not mark distinct senses? Raindrops felt on my head are like a car coming into my field of vision; touching my head is like peering out of the window to catch sight of the car just disappearing up the road.

I would respond that the difference is that the judgement "My legs are crossed" is always – barring exceptional circumstances such as proprioceptive deficit – redundant. In contrast, even when seeing and feeling do

not actually involve action, there is always a *possible* action of looking or touching. One could gain further knowledge of the surface of the sphere through active exploratory touch, or turn to look at the car as it disappears up the road. There is no parallel here with proprioceptive touch or with proprioception in general. I avoid objects in my path automatically, just as I uncross my legs automatically in standing up, but it would be wrong to say that I "just know" that there is a large tree in front of me, or I "just know" where the piano is when I can remember where it is. I know that it is in front of me because I can see it or because I can remember where it is. Knowledge of the immediate environment is not immediate knowledge.

There are further reasons for denying that proprioception is one of the senses. If it were, feedback about error would be required, yet the only sense-specific feedback or correction comes from the five senses. There are more fundamental ways in which proprioception and the senses are interdependent, for sensory orientation *assumes* proprioception – in particular visual proprioception or visual kinaesthesis. One cannot orient one's sense organs without knowing whether one is moving or stationary oneself – information derived from proprioception. This insight is associated with the psychologist J. J. Gibson, who proposed the interdependence of perception and proprioception, and made familiar the concept of visual proprioception. We now see how his views offer support to the non-perceptual account of proprioception that I am advocating.

4.4 Gibsonian and other considerations against the perceptual model

Gibson's work places a particular interpretation on the denial that proprioception is one of the senses, and while endorsing aspects of his account, I need to distinguish my own position from his. We saw earlier that unlike Sherrington, he regards proprioception as a component of the functioning of all the perceptual systems. He claims that proprioception and perception are interdependent; they are distinguished in terms of their function, and not by the receptors, sensory nerves, or sensations involved. For him, a perceptual system is not to be identified with a sensory modality, a specific channel of input transmitted by specialised sensory nerves to the brain for processing. According to Gibson, perceptual systems are the means by which we actively pick up meaningful information about the environment: "All the perceptual systems are propriosensitive as well as exterosensitive, for they all

provide information in their various ways about the observers' activities".[29] These two functions are interdependent, he argues: "an environment implies something that is surrounded, and therefore awareness of the environment implies an awareness of the body existing in the environment. Equally, an awareness of the body entails some feeling of its relation to the surroundings".[30]

Gibson was particularly interested in what he called visual proprioception. He regards vision as proprioceptive in that like the muscle-joint-skin system and the inner ear, it "obtains information about *both* the environment and the self".[31] For Gibson, visual proprioception, the least automatic and "highest" variety of proprioception, is central to action-guidance, especially in any new task. The movement sensitivity of the visual system dominates that of the muscular and articular systems in manipulation and locomotion: "we see where we are going and the layout of the environment through which we are going at the same time ... vision is kinaesthetic in that it registers movements of the body just as much as does the muscle joint system and the inner ear system".[32] Visual kinaesthesis explains perception of passive as well as active movement – passengers sitting in a car perceive that it is they who are (passively) moving, and not the trees, buildings, and road rushing past.[33] Other senses also have a propriospecific component:

> information about the self is multiple, and ... all kinds are picked up concurrently ... An individual not only sees himself, he hears his footsteps and his voice, he touches the floor and his tools, and when he touches his own skin he feels both his hand and his skin at the same time. He feels his head turning, his muscles flexing and his joints bending. He has his own aches, the pressures of his own clothing, the look of his own eyeglasses – in fact, he lives within his own skin.[34]

---

[29] Gibson (1979), p. 115.
[30] Gibson (1968).
[31] Gibson (1979), p. 183.
[32] Gibson (1968), p. 36; (1979), p. 185. Gibson's interest in visual proprioception grew out of his wartime experiments in training pilots. He found that while the plane tilts, the pilot's horizon or "straight ahead" does not slope but remains at eye level; while non-visual proprioception is often unreliable, the pilot's visual world remains stable. See Reed (1988), p. 78.
[33] Visual proprioception is discussed by Butterworth (1995), pp. 93–6. It plays an important role in the critique of self-presentation in the next chapter.
[34] Gibson (1979), p. 115.

Gibson believes that all these varieties of proprioception are distinguished from perception, which they nevertheless accompany through their function of "self-sensitivity" or "egoreceptivity".

The model of knowledge of bodily position as continuous, based on constant feedback from touch and sight and other sensations, is a persuasive one. Gibson's claim that proprioception is the product of various senses implies a convincing rejection of the perceptual model. He is right to suggest that, while some proprioceptive systems – such as those concerning general fatigue and nutrition – provide information solely about the body, others provide information about both the subject and the environment. The vestibular system, for instance, is concerned with bodily balance and hence with the relation between the body and the environment, and similarly with touch. However, even Gibson's supporters acknowledge defects in his line of argument.

For instance, Bermúdez in *The Paradox of Self-Consciousness* defends a Gibsonian account of proprioception.[35] He maintains that perceptual experience not only provides information about the external world, but is inextricably combined with self-specifying information without which the former would be of little use; he argues that Gibson's ecological approach to perception shows that perceptual experience is a source of first-person non-conceptual contents.[36] However, Bermúdez goes on to argue that Gibson's insights into the structure of visual perception are partly vitiated by his neglect of proprioceptive bodily awareness in distinguishing passive from active self-movement; visual proprioception enables the subject to distinguish self-movement from movement of the environment, but cannot tell them whether or not they are moving under their own steam.

This highlights a key issue, I believe. It has been a common complaint that Gibson's ecological theory of perception neglects the subject's role in organising their perceptions.[37] But the problem is more specific. In deflating the distinction between proprioception and exteroception, Gibson undermines the status of self-consciousness. If proprioception is interdependent with exteroception, one must also recognise that it is distinctively self-conscious, as shown by its exhibition of IEM. There is thus a tension between Gibson's picture of proprioception as continuous

[35] Bermúdez (1998).
[36] Bermúdez (1998), p. 114.
[37] As Reed (1988) argues, p. 303. Reed is, I think, too sanguine in suggesting that Gibson has a theory of self-consciousness in any very developed philosophical sense.

with other perceptual modalities, and Bermúdez's claims that proprioception counts as a "genuine form of self-consciousness", and that there is a close connection between IEM and the essence of first-person judgements.[38] The Gibsonian picture unduly deflates self-consciousness. It fails adequately to acknowledge the IEM of proprioception, since it does not allow its distinctive ground. Gibson's account requires development in order to reject the perceptual model of proprioception while still acknowledging self-consciousness.

I conclude the case against the perceptual model by arguing that it fails to acknowledge a crucial ambiguity in the concept of feeling. The denial that proprioceptive knowledge is immediate – that I do not have to do anything in order to have it, and "just know", for instance, that my legs are crossed – arises, I will argue, from a misunderstanding of the concept of feeling that is characteristic of the image theory and perceptual model. Proponents of the perceptual model claim that in shifting my attention from the object being touched to the sensations that I enjoy while touching it, I am simply moving my attention from objects that lie outside one of my bodily boundaries, for instance the surface of a hand, to what is occurring at or beneath that boundary. Hence according to Michael Martin our bodily experiences "have as part of their phenomenological content that the region felt falls within one's body".[39]

Since content is essentially propositional, I believe, the claim should not be presented in terms of "phenomenological content". Martin's assumption that the same sense of "feeling" is involved in both exploratory and proprioceptive touch should in any case be rejected. Exploratory touch aside, I do not "feel" my body in the same way that I feel other "regions". One can become aware of specific bodily feelings, but not in the way that an image or perceptual account implies. Feelings of pressure, proprioceptive touch, muscle tension, or skin stretching do not involve feeling in the perceptual sense exhibited in exploratory touch.

---

[38] Bermúdez (1998), p. 144 and elsewhere. Bermúdez's discussion of proprioceptive content is fertile and suggests many questions for further discussion, especially concerning the dual criteria for location, and the concept of a hinge (1998, pp. 154–61).

[39] Martin (1995), pp. 270, 273. He denies that bodily awareness is a matter of "non-perceptual states immediately caused by action on the body". One target here may be the implausible Cartesian introspectionist view that bodily sensations are "non-perceptual sensory experiences"; another might be the direct knowledge position. A perceptual view is pervasive in Bermúdez et al. (eds) (1995), and in Bermúdez (1998), for instance p. 135.

Some remarks of Wittgenstein show why this is. Normally – that is, when the subject is not suffering from proprioceptive deficit – the question "How did you know that your legs were crossed?" is redundant. The only sensible answer to the normally otiose question "How do you know that your legs are crossed?", is "I just know". Wittgenstein acknowledges this, while apparently leaving space for a non-perceptual sense of "feel":

> My lower arm is now lying horizontally and I should like to say I feel that; but not as if I had a feeling that always goes with this position ... - rather as if the "bodily feeling" of the arm were arranged or distributed horizontally, as e.g., a film of damp or of fine dust on the surface of my arm is distributed like that in space. So it isn't really as if I felt the position of my arm, but rather as if I felt my *arm*, and the feeling had such and such a *position*. But that only means: I simply *know* how it is lying – without knowing it *because* ... As I also know where I feel pain – but don't know it *because* ... [40]

Wittgenstein here assimilates knowledge of bodily position with knowledge of the location of pain in a quite different way to Martin.

His remarks distinguish three senses of "I felt my arm":

(1) Perceptual/exploratory: I grab it with the other hand.
(2) Non-perceptual bodily feeling: my arm stings.
(3) Proprioceptive: I just know how it is lying and that it is sensitive.

Wittgenstein's suggestion is that if, for example, I have a stinging sensation all over the surface of my lower arm – caused, say, by dilute acid or stinging nettles – then it is true that I feel my arm, and that feeling has such and such a position. But his assimilation of bodily position with knowledge of pain location constitutes a denial of the perceptual model also, for the claim that it is "as if the 'bodily feeling' of the arm were ... distributed horizontally" suggests that I am not perceiving an object at all. His implication might be that "I felt my arm" in the proprioceptive sense can only mean "My arm wasn't asleep, frozen, or anaesthetised, and I just know how it is lying". It suggests that the subject of

[40] Wittgenstein (1980), vol. I, para. 786; the idea is also expressed in paras. 784–5. See also Wittgenstein's discussion in his (1958) Part II, pp. 185–6; and Merleau-Ponty (1962), p. 93, on the localisation of pain.

proprioception is not simultaneously an object of perception, and that proponents of the perceptual model are mistaken in claiming that it is.

The fact that proprioception yields knowledge is the strongest reason for supporting a perceptual model. But if one holds, with Wittgenstein, that knowledge implies the logical possibility of doubt, then the existence of proprioceptive knowledge rests on the fairly remote possibility of error of such judgements as "My arms are folded" or "My legs are crossed".[41] Though the judgement is not infallible, the subject's knowledge of their posture and position is normally so reliable as to be almost certain. Except in pathological cases, errors involve inattention, and are limited to rather complex situations – intertwining one's hands for instance – or to errors of detail that result from imprecise proprioceptive knowledge discussed earlier.[42] In contrast, except in unusual cases of phantom-limb phenomena, it is hard to see how I could be mistaken that my legs are crossed. (Such phenomena often involve pain in the missing limb, but patients also (mis-)report its position too, for instance claiming that it feels as if it is buckled under them.[43])

To return briefly to the epistemic congruence of proprioception and memory: memory is also a faculty which yields direct, immediate, and reliable knowledge, and is not a mode of perception. However, the kind of knowledge that proprioceptive judgements express, if it is knowledge, is of a curiously muted variety; the knowledge-claims are oddly unconvincing. Nonetheless I do not propose to deny that proprioception mostly does yield knowledge, otherwise an unconvincing assimilation of judgements of bodily position to avowals of sensation results. Proprioception has an ambiguous status, therefore. It seems to involve knowledge of an object – one's body – but knowledge that is groundless, at least in what are usually referred to as "internalist" terms. This ambivalent status matches its intermediate position on a continuum

---

[41] Wittgenstein's view that knowledge implies the logical possibility of doubt is defended in Hamilton (2013 forthcoming).

[42] This reliability so impressed Danto that in "The Body–Body Problem" and other articles in his (1999a), he several times suggests that immediate bodily knowledge might be indubitable. (Like Anscombe he does not use the term "proprioception".)

[43] A more extreme example is A. R Luria's patient Zasetsky, "The Man with a Shattered World", who during World War II received massive brain damage : "Sometimes when I'm sitting down I suddenly feel as though my head is the size of a table…When I close my eyes, I'm not even sure where my right leg is; for some reason I used to think (even sensed) it was somewhere above my shoulder" (Luria 1972, pp. 42–3).

between perceptual knowledge from the five senses, and awareness of sensations such as pain.

This concludes the main part of the critique of the sensory model of proprioception.[44] Before turning to the issue of IEM and proprioception, central to the line of argument in Chapters 4 and 5, we must consider the connection between bodily awareness and action.

4.5 Proprioception and agency

IEM is exhibited by *basic action*. On Danto's well-known definition, raising one's arm is normally a basic action: I just do it, and do not do it by doing something else. "I raised my arm" is not (normally) intentional under the description "flexing my biceps". Unusually, I could do it by lifting it with my other arm, and this would be a non-basic action. More commonly, I can open a door that is already ajar by pushing it with my hand, kicking it, or shoulder-charging it; opening the door is a non-basic action which I do by performing one of these basic actions.[45]

We should first consider the parallel between the perceptual model of proprioception and the model of basic action associated with it, what I have termed *the control model*. This model constitutes one element of the Cartesian residue in Husserl's discussion considered earlier. Husserl describes "my body" as "animated flesh" (and bones), and refers to the "psychological I" as that aspect of the lived Body which animates it. His claim that the lived Body is unique in "being moved 'spontaneously' or 'freely' by the will of the Ego" only serves to qualify the Cartesian mechanistic model of bodily "control".[46] The control model is inadequate first because it implies such unconvincing causal claims as "When I raise my arm (basically), I cause it to rise". If I tell someone "I caused my arm to rise", and they reply in puzzlement "How did you do that?", it would be a feeble joke to say "I raised it". I cause my arm to rise only when I move it non-basically: by doing something else, such as pressing a switch which causes a sling to raise it, or maybe by some parapsychological feat that causes my arm to levitate.

---

[44] There are further arguments against it in Chapter 5, which show that lack of proprioception is a far more radical deficiency than blindness.

[45] Danto (1965). The notion is not unproblematic. Is signing my name a basic action, which I do by moving my hands while holding a pen? Or playing a practised routine on the piano, by pressing one's fingers down in turn? Perhaps it is an atomistic, traditionally empiricist notion, as opposed to a gestalt empiricist one. Pursuit of these issues must await a later occasion.

[46] Husserl (1989), p. 167.

The Cartesian picture – "Cartesian", to reiterate, in the textbook sense – is inadequate also because I do not ordinarily control or exert control over my body. Only in unusual circumstances could one speak of such control – as perhaps when a Parkinson's sufferer controls one trembling hand with the other, better functioning one. In this way also, perhaps, a sufferer from Tourette's syndrome is able, by great willpower, to suppress their tics, only for them to explode with violence later. It would be some improvement on the control model to say that, in the normal case, I act through or by means of my body. But even this claim still suggests Cartesian mechanism. Action is essentially a bodily phenomenon, though as earlier comments suggested, this does not mean that it is purely physical. Although his predilection for the metaphor of control renders Husserl's account unsatisfactory, his claim that the lived Body is unique in "being moved 'spontaneously' or 'freely' by the will of the Ego" may be an attempt to capture the more fundamental of two senses of "move", the sense in which one moves one's body basically.

Basic movement, analogous to basic action, is characterised more accurately by Merleau-Ponty, though the control model is still implicit: "I move external objects with the aid of my body, which takes hold of them in one place and shifts them to another. But my body itself I move directly, I do not find it at one point of objective space and transfer it to another…".[47] I move my arm basically when I move it not by doing something else, for instance by picking it up with my other hand. Thus I mostly move my body basically, though I can move it non-basically. In the case of non-basic action, errors of self-identification are possible. It could be that coincidentally with my moving someone else's arm, thinking it is mine, someone else moves my arm, and so it seems as if I am moving my arm in the non-basic sense. Or I could be paralysed, initially without realising it, and think that I am moving my arm in a non-basic way when in fact someone else is. Self-ascriptions of basic actions, in contrast, exhibit IEM.[48]

There is a further sense of control, distinct from the still relatively normal sense in which a Parkinson's patient tries to control a trembling hand, which brings out the intimate connection between proprioception and action: the sense in which patients with severe proprioceptive deficit may be said to "control" their body. For them action is transformed, and

[47] Merleau-Ponty (1962), p. 94.

[48] A subtler difference from the present account is that Husserl, like Merleau-Ponty, treats intentional attributions as made to the lived Body, whereas on the present account they are ascribed to persons, an issue pursued in the next chapter.

the body itself becomes almost a prosthetic device. Cole, in his case study of "G. L.", writes that she "occasionally talks of her body as being a machine on which she imposes commands ... A more accurate description may be that she uses her body as a tool, a passive instrument that can be used to move and to interact with her environment".[49] This kind of agency is conceptually as well as practically impoverished.

The example brings us to a second and larger issue: how proprioception underlies the possibility of action, and so is essential to self-consciousness. A subject who suffers from severe proprioceptive deficit has restricted capacity for action; a subject with total proprioceptive deficit could not act at all. If Ian Waterman did not have residual proprioception from the neck upwards, he could not have learned to walk again. There has to be a control centre with some proprioceptive capacity, even if the body it controls is prosthetic. The faculties which Ian Waterman uses to get his largely prosthetic body to "act", highlight how agency requires sensory perception, which in turns requires proprioception. In order to act, a creature has to orient some sensory modality or modalities – it has to look, or feel, or orient its ears. This is impossible without proprioception: for instance, the capacity of looking involves the capacity of moving the head.[50] These considerations yield a further argument for concluding that proprioception is not one of the senses; it underlies the possibility of perception, and so cannot be a variety of it. Proprioception is essential to self-consciousness, most fundamentally because it underlies the possibility of action.

4.6 IEM and q-proprioception

We now turn to the central outstanding issue of the discussion of the body and self-consciousness: IEM of bodily self-ascription. This was neglected by Wittgenstein in his contrast of "I"-as-subject and "I"-as-object in *The Blue Book*, which included only psychological self-ascriptions under the first heading. This was an oversight, as proprioceptive judgements do exhibit IEM – a fact that, as Evans noted, constitutes an antidote to Cartesianism.[51] It is true that judgement is not central to proprioception as it is to personal memory – though even there, I have argued, its

[49] Cole (1995) in Bermúdez et al. (eds) (1995), p. 261.

[50] A stress on the active nature of seeing, and the need for feedback and exploration to be genuinely embodied, is found also in recent post-Gibsonian work in sensorimotor contingency theory, such as Noe (2006).

[51] Evans (1982), p. 224 and elsewhere.

importance can be overstated. Proprioceptive knowledge is expressed principally in bodily movement and action; proprioception is a practical capacity only marginally concerned with propositional knowledge. But since the IEM claim is essentially a negative one – the senselessness of a certain kind of judgement or doubt – the relative unimportance of judgement does not invalidate the parallel with memory.

I will argue that immediate judgements of bodily position, posture, and movement – those which express proprioceptive knowledge, unlike judgements about one's body based on visual perception or touch – are IEM. This claim takes the same form as in the case of memory. If I judge immediately that my legs are crossed, and then come reasonably to doubt this, it will be senseless for me to cite the original justification as a reason for believing that nonetheless *someone's* legs are crossed. That is, it will make no sense for me to say "Well, I at least just know that someone's legs are crossed". "I just know" functions like the continuous-verb locution "I remember o-ing" in personal memory. (Even to talk of justification sounds forced, since, I will argue, in the normal case I just know that my legs are crossed.) In contrast, when in unusual circumstances the subject judges, on the basis of vision, that her legs are crossed, there is no IEM guarantee.

IEM applies also to judgements of limb-ownership, and judgements concerning basic actions. Thus it makes no sense to ask "I moved someone's arm (basically), but was it my arm which I moved?" or "Someone moved my arm (basically), but was it myself who moved it?" In the case of a non-basic action, both questions could make sense, though the first case would require that there was no proprioceptive feedback from the arm that I moved. The second case may be illustrated by the following unlikely scenario: a row of patients each has their arm in a sling, and this can be moved by pressing one of an array of switches. (The bionic arm discussed below would need careful treatment to sustain these claims.)

At this point, however, I focus on the IEM of proprioceptive judgement. In contrast to the case of memory, the formulation of proprioceptive IEM is slightly problematic, because except where proprioception is temporarily or permanently impaired, it is hard to think of a situation where I could doubt that my legs are crossed. Perhaps there is an "after-legs-crossed" sensation, analogous to the "after-hat" sensation of tightness after wearing a hat, which causes someone in the dark to be unsure whether their legs are crossed. But this seems unlikely. Since the very possibility of doubt is put into question, one cannot get to the point of imagining a doubt on certain grounds in order then to rule it out. Proprioceptive judgements thus approach the IEM status of avowals.

Connectedly, there is a lack of credible real-life counter-examples to the IEM of "My legs are crossed" – no equivalent, within normal experience, of the information-garbling discussed in the case of memory.

The lack of real-life counter-examples to proprioceptive IEM contrasts with the case of memory. As we saw in Chapter 2, there are tokens of the type-judgement, "I swallowed a ball bearing when I was a child" based on personal memory, and ones based on factual memory or testimony. Judgement-types concerning bodily position and movement also have two kinds of token: those that express proprioceptive knowledge of bodily position, and those that are made on the basis of visual perception or exploratory touch. In contrast with memory, however, the subject cannot be mistaken concerning the basis of her judgement. While she can say "I swallowed a ball bearing when I was a child, though I'm not sure whether I remember this, or know it because I was told it", it would be absurd for her to say "My legs are crossed, but I'm not sure whether I just know this, or know it because I can see that they are crossed".

Blindsight might seem to make the confusion just intelligible, since here, arguably, the subject knows what is in front of them while denying that they see it. However, the phenomenon is misinterpreted and philosophically overrated. As we saw in Chapter 1, when verbal and other behavioural criteria diverge drastically, there is no determinate answer concerning whether the subject sees or not. If my lower limbs have been locally anaesthetised all over, then just possibly I could look into a mirror and think that they are crossed when someone else's are, before rapidly correcting myself. But while Ian Waterman relies on constant visual monitoring, no one with a functioning proprioceptive capacity could seriously "monitor" their body in this way.

As with q-memory, the phenomenon must be carefully defined. Q-proprioception should be a peculiar kind of proprioception, involving an information-link between the subject and someone else's body, such that the subject allegedly registers information from it. Armstrong, for instance, writes: "We can conceive of being directly hooked-up, say by a transmission of waves in some medium, to the body of another. In such a case we might become aware e.g. of the movements of another's limbs, in much the same sort of way that we become aware of the motion of our own limbs".[52] "Alteroception" might be preferred to the oxymoron "q-proprioception", since "proprioception" is a contraction

[52] Armstrong and Malcolm (1984), p. 113. Similar thought-experiments are found in Evans (1982), p. 221, and Cassam (1997a), p. 63.

of "proprioperception" or "self-perception". But I continue to use "q-proprioception" because of the symmetry with q-memory.

Q-proprioception may be defined, analogously to q-memory, as follows: a subject's judgement "My legs are crossed" expresses q-proprioceptive knowledge if and only if

(i) he apparently has direct and immediate knowledge of limb-position;
(ii) that apparent knowledge involves information from the body of someone who is not necessarily the person who appears to have the knowledge;
(iii) the subject did not receive the information in a non-proprioceptive or extraneous way, that is, q-proprioception is a peculiar kind of proprioception.[53]

Compare q-proprioception with the science-fictional scenario of *q-pain*. In post-amputation phantom-limb experience, the phenomenal location of the sensation is where the limb once was. Q-pain, so-called, extrapolates this situation, so that the subject feels a pain that is phenomenally located in another person's body. In contrast to q-proprioception, however, there should be no inclination to deny that the pain is mine, and so this is not properly a q-phenomenon at all.[54]

As we saw, there is no real-life information-garbling in the proprioceptive case. But, as with memory, there is a contrast between mechanistic and non-mechanistic cases of q-proprioception. Science-fictional scenarios envisage an information-link between the subject and someone else's body, such that the subject allegedly registers information from it. The apparent intelligibility of this mechanism – the identification of proprioceptive information with nerve impulses – makes q-proprioception seem more plausible than q-memory. However, it is subject to the same fundamental critique of the impersonal concept of information, and the associated postulation of a determinate information-link existing prior to and independent of a way of knowing. Nerve impulses cannot be said to transmit information concerning bodily

[53] Again a distinction between strong and weak q-proprioception may be proposed, with similar arguments against it as in the case of memory (see Chapter 2).

[54] The distressing symptoms of thalamic syndrome include "generalised pain". Specific neural pathways are blocked and only the diffuse pathways operate, and subjects typically report that the pain is "everywhere and nowhere".

states – information in the ordinary, not the impersonal, technical sense – until it is known whether the subject comes to know anything as a result. Proponents of q-proprioception and q-memory move illicitly from an impersonal, knowledge-free concept of information, to an information-link in a richer, epistemic sense. One cannot assume, with Evans and Cassam, that the person immediately registers the information; whether the brain does so depends on the person's reaction.

In fact q-proprioception is not easy to characterise on the model assumed by Armstrong and others. There are a range of possible scenarios, the most plausible where information from the neurally attached alien subject overrides the subject's own bodily awareness – as when my arm is by my side and the other subject's is in the air, and I judge "My arm is in the air". Unless external proprioceptive information overrides internal – and I avoid looking at my arm – my judgement would be confused. The subject would have to have a proprioceptive deficit, and be linked to the body of another with proprioceptive deficit, who was also paralysed. This is hard to conceive without an efferent nerve-connection operating in tandem with an afferent one, enabling the first subject to raise the other's arm, while the latter is locally paralysed, to avoid a "conflict of wills". However, increasing the connections in this way makes the two bodies seem less distinct – hence what may be termed the *two mere bodies, one lived Body dilemma*. A plethora of neural connections makes clear proprioceptive knowledge more likely, but also suggests that the two mere bodies constitute one lived Body in the Husserlian sense, which rules out q-proprioception. But a small number of nerve connections, in contrast, fails to guarantee clear proprioceptive knowledge.

It may be responded that two distinct subjects could "share" a hand, without their two mere bodies constituting one lived Body. In this scenario, however, ownership of the hand itself is in doubt. Assume that I have IEM-exhibiting knowledge of, and can move basically, a hand which is part of the material unity constituted by your torso, and so forth. On the materialist criterion to be criticised in Chapter 5, which says that a body part belongs to a subject if and only if it is part of the material unity which they are, the hand is unquestionably yours. But this claim is doubtful; if another agent can move the hand basically, it is at least unclear whose hand it really is. (Though one may also wonder how my movement of the hand could constitute a basic action.) So the alien-limb scenario is the basis for an important line of objection to q-proprioception.[55]

[55] Material unity is not a straightforward notion, as we see in the next chapter.

The preceding discussion of q-proprioception assumed a specified mechanism for the information-link. But, as in the case of q-memory, there is an alternative scenario where no neural connection is postulated, and no prior determinate information-link presupposed. The subject is simply inclined to say that their own legs are crossed, an inclination which they have learned is correlated with someone else's legs being crossed. There is a parallel with telepathy, and one might question whether the knowledge is proprioception-like. Preceding arguments – the two mere bodies, one lived Body argument, the argument from limb-ownership, and the objection against q-memory that the knowledge is inferential – still apply.[56] (The issue is pursued in Chapter 5.)

Thus far I have been predicting how subject X would behave in science-fictional scenario Y – assuming that there is a fact of the matter concerning questions such as "Whose limb?" But the effect of "external" nerve connections on an integrated teleological system is unpredictable; could the subject make coherent sense of the new information? To ask whether alleged immediate knowledge of an alien limb could cohere with the subject's total proprioceptive knowledge is like asking whether q-memory could be integrated with the holistic, autobiographical structure of personal memory. Despite appearances, therefore, q-proprioceiving the body of someone else may not be more intelligible than q-remembering their past; indeed, because of immediate conflict with other sensory input, especially visual, it might be more unsettling. But these predictions are speculative, as are the philosophical conclusions made on their basis. Wittgensteinian restrictions on determinate concept-application should make one question whether there is a fact of the matter concerning limb-ownership, in science-fictional scenarios.

There is therefore a clash of intuitions. A pure self-conscious knowledge account suggests that the alien limb belongs to the person who experiences it and moves it basically. However, while self-conscious knowledge normally coheres with material unity, it does not do so in the alien-limb scenario. The dissonance between self-conscious knowledge and material unity leaves us at a loss what to say. The self-conscious

[56] Evans uses the distinction between registering information and gaining knowledge in his criticisms both of q-proprioception and q-memory (Evans 1982, p. 221). It is also implicit in Martin's view that "feeling a pain in someone else's hand" counts as an illusion or hallucination, and that the existence of a causal or informational link between body parts and perceiver does not suffice for perception of those parts (1995, p. 283).

knowledge intuition rules out the materialist view that despite my having IEM-exhibiting knowledge of it, the alien limb clearly belongs to a distinct subject, but it does not imply that the limb clearly belongs to me.

The account defended here denies that proprioceptive knowledge is a conceptual superimposition on pre-identified and individuated living Bodies; rather, it is part of the material from which the latter concepts are formed. Merleau-Ponty's postulation of a "body-subject" implies that central varieties of bodily self-ascription are immediate and IEM, and suggests that proprioception is not simply a mode of awareness of events that happen to occur within the body's boundaries. Proprioception does not just happen to yield knowledge only of one subject, as seeing oneself in the mirror does, and should not be regarded as a capacity for tracking a fundamentally material entity. The self-conscious knowledge elucidation of "my body" as "the one of which I have self-conscious (IEM-exhibiting) proprioceptive knowledge and which I can move basically" suggests, rather, that proprioception and bodily identity form a conceptual holism or circle of inter-defined concepts. Just as the concepts of personal identity and self-conscious, IEM-exhibiting capacities – notably memory – interlock, so do the concepts of bodily identity and self-conscious, IEM-exhibiting bodily awareness and movement. The next chapter offers a defence of the resulting qualified self-conscious knowledge account.

5
Proprioception and Self-Consciousness (2): Self-Conscious Knowledge and the Rejection of Self-Presentation

This chapter has two principal aims. The first – in Sections 1 and 2, following the non-sensory treatment of proprioception in Chapter 4 – is to develop and defend a *self-conscious knowledge account of bodily identity*, in opposition to the almost universal position of *materialism about bodily identity*, which regards the body as fundamentally a material entity. In the previous chapter, a self-conscious knowledge account was outlined in connection with the alien-hand scenario. This account says that "my body" is the body of which – when conscious – I have self-conscious (proprioceptive) knowledge, and which I can move in a basic sense, that is, not by doing something else. The account is a development of the Lockean view that to experience a limb as mine – to feel it when it is touched, to be conscious of it as hot or cold and as having other "affections", to have sympathy and concern for it – is necessary and sufficient for it to be mine.

The second aim – in Sections 3 to 6 – is from the same basis of a self-conscious knowledge account of the body, to offer a critique of a somewhat distinct kind of materialism. This is Cassam's version of what he terms materialism about self-consciousness, and which I call the *self-presentation thesis*, which says that we can and must be presented to ourselves, qua subjects, as physical objects among physical objects.[1] My

[1] Other writers also defend it, notably Bermúdez, whose (1998) argues that it is only in virtue of the sense of touch that the body is experienced as a solid and bounded entity in the world.

critique argues that the truth in self-presentation is just that proprioceptive judgements are IEM, and that IEM is integral to self-consciousness. Any further content to self-presentation is refuted by a *contrast argument*, which cites the lack of any intelligible contrast to the experience of self-presentation. The argument turns out to depend on the possibility of a subject who has a total and congenital proprioceptive deficit. The contrasting experience – lack of experience of embodiment – must be had by a self-conscious subject, and it is doubtful that this subject is self-conscious, or even has experience.

The chapter concludes by returning to the opposition between the unitary self and the unity of consciousness, relating proprioception and the body to earlier discussion of that opposition.

5.1 Self-conscious knowledge versus materialism concerning bodily identity

The previous chapter described the almost universal assumption that the body is a purely material and epistemically outer entity, whose states, in contrast to those of the mind, are not known immediately and transparently. I called this position *materialism about the body*, and – to reiterate – almost all Analytic philosophers, except those influenced by the Phenomenological tradition, subscribe to it. The associated position *materialism about bodily identity* holds that a body-part is mine if and only if it is part of the material unity that I am. The present chapter defends, against these positions, the truth in a *self-conscious knowledge account*, which says that "my body" is the body of which – when conscious – I have self-conscious (proprioceptive) knowledge, and which I can move in a basic sense.[2] Roughly speaking, on this account, X is a body-part of mine if and only if I have IEM-exhibiting knowledge of it, and can move it basically. "My body is the body of which ..." is an elucidation of the concept of my body, rather than a criterion of identity, since, as I will argue, there is no conceivable use for such a criterion.

In discussing this dense and complex issue, it helps to begin historically. Locke interestingly anticipates the self-conscious knowledge account,

[2] The phrase "my body" emphasises the subject's knowledge of and concern for his own (living) body. Carnap, influenced by Husserl's concept of the lived Body, used it in this sense in the *Aufbau* (Carnap 1967, Sections 129–31). Danto contrasts "our bodies" with "mere bodies" (1999c, p. 64).

and the Phenomenological tradition in general, in his comments on bodily unity and identity:

> [all] particles [of our very bodies], whilst vitally united to this same thinking conscious self, so that we feel when they are touched, and are affected by, and conscious of good or harm that happens to them, are a part of our*selves*; i.e. of our thinking conscious *self*. Thus, the limbs of his body are to every one a part of *himself*; he sympathizes and is concerned for them. Cut off a hand, and thereby separate it from that consciousness he had of its heat, cold, and other affections, and it is then no longer a part of that which is *himself*, any more than the remotest part of matter.[3]

These remarks suggest an *experiential condition* for ownership of a body-part, namely, that I experience it as mine: feel it when it is touched, am conscious of it as hot or cold and as having other "affections", and feel "sympathy" and "concern" for it. This condition, Locke suggests, is necessary and sufficient for limb-ownership.

The condition is not sufficient as it stands, however, and requires development. First, it should be stated more precisely in terms of bodily self-knowledge that exhibits IEM: I must have immediate, proprioceptive, IEM-exhibiting knowledge of the limb's position and movement, and must be able to move it basically.[4] We need to re-trace at this point the discussion in Chapter 4 concerning IEM-exhibiting knowledge. Distinctively self-conscious knowledge, which exhibits IEM, is manifested in judgements of posture, orientation, intention, and action. If I judge immediately that my legs are crossed, and then come reasonably to doubt that they are, it will be senseless for me to cite the original justification as a reason for believing that nonetheless *someone's* legs are crossed; it will make no sense for me to say "Well, I at least just know that someone's legs are crossed". In contrast, judgements on the basis of vision that my legs are crossed are not IEM. We should note here that IEM applies also to judgements of limb-ownership and judgements concerning *basic actions*; it makes no sense to ask "I moved someone's arm (basically), but was it my arm which I moved?", or "Someone moved my arm (basically), but was it myself who moved it?" For a non-basic

---

[3] Locke (1975), Book II, Chapter 27, para. 11. The term "idealism" is used in this context by Cassam (1997a), p. 63.

[4] To reiterate, I move my arm basically when I move it not by doing something else, for instance by lifting it with my other hand.

action, both questions could make sense, though the first would require lack of proprioceptive feedback from the arm that I moved. The second is illustrated by a row of patients each with their arm in a sling, which could be moved by pressing one of an array of switches.[5]

A Lockean account, therefore, implies the following biconditional for the experiential condition:

> (1) X is a body-part of mine *if and only if* I have IEM-exhibiting knowledge of it, and can move it basically.

Clearly this is still inadequate. Human subjects are so constituted that there are parts of one's body – such as the brain – of which one does not have proprioceptive, IEM-exhibiting knowledge, and cannot move basically. So we need to develop (1) into a bipartite account. In doing so, there need still be no appeal to material unity:

> (2) X is a body-part of mine *if and only if* either (i) I have IEM-exhibiting knowledge of it, and can move it basically; or (ii) it is part of a body-part that satisfies condition (i).

This version of the Lockean biconditional rests simply on the transitivity of parthood, which is a basic principle of classical mereology, namely: if A is a part of B, and B is a part of C, then A is a part of C.[6] Thus, for instance, the brain is a body-part in virtue of being part of a part that satisfies condition (i), namely, the head.

Biconditional (2) is plausible, but one must still decide between interpretations of the right-hand side of the biconditional, as *extension-determining* or *extension-reflecting*. A Lockean account treats the IEM-exhibiting capacities as criteria of bodily identity; it holds that IEM-exhibiting knowledge *determines* whose body-part X is, rather than merely tracking an extension that is independently (materially) determined.[7] In other words, it holds that it is the subject's experience that decides who owns the body-part, rather than treating that experience as a mere symptom of an underlying material criterion, that they are materially united with the body-part.

---

[5] The example of the bionic arm discussed below would need careful treatment to sustain these claims.

[6] Varzi (2006) defends this principle against recent critics.

[7] The contrast between extension-determining and extension-reflecting judgem-judgements is discussed in Wright (1992), pp. 108ff.

When defined in this way, the Lockean position might be described – as Cassam does – as "idealist". However, that would be misleading.[8] Clearly the Lockean position allows that subjects are embodied, and that the limb is a material part of a material entity. It is only who *owns* the limb that depends on how it is experienced. The Lockean view is therefore idealist in a very diluted sense: it is an "idealism of ownership" only.

Even this residual "idealism" is unacceptable, however. Thus my account parts company with the Lockean position, offering instead a *no-priority interpretation* of the biconditional. The distinctive treatment that results requires careful elucidation. It involves a rejection of the dichotomy of extension-reflecting and extension-determining; neither the LHS or the RHS has priority. A no-priority interpretation expresses what I term a *conceptual holism* between my body and body-parts on the one hand, and proprioceptive knowledge and basic movement on the other. As we saw in Chapter 1, conceptual holism consists in an equivalence and interdependence between the concepts concerned, such that neither is more basic than the other. That is, a definition or understanding of concept X makes essential reference to that of concept Y, and vice versa – there is a relation of mutual presupposition between them. So just as the concepts of personal identity and memory interlock – which, as we saw in Chapter 3, constitutes the truth in the memory criterion of personal identity – so do the concepts of bodily identity and self-conscious, IEM-exhibiting bodily awareness and movement. The disanalogy is that the latter holism is synchronic, while that concerning memory is diachronic.[9]

The account of bodily identity that I am defending, therefore, is not implausibly idealist; it recognises the significance of *material unity*. This is not a simple notion. As Ayers notes, it does not consist simply in particles "merely coming to lie adjacent to one another, like marbles in a bottle": "the unity which life helps to constitute *is* material unity, not a rival to it. The animal and the bullet [lodged in it] are not, in the sense required, *materially* united".[10] It is closer to a biological concept, therefore; "organic unity" is an alternative description, since an organism is defined solely by its material unity.

---

[8] Cassam (1997a), p. 63.

[9] Bodily identity over time would involve memory also, and would be subservient to personal identity on the principle "same person, same body", and vice versa.

[10] Ayers (1991) vol. 2, pp. 229–38; quotation, p. 232.

The self-conscious knowledge account that I offer acknowledges that my body is *normally* a material unity, while denying that material unity affords a sufficient criterion of bodily identity, or identity of body-parts. This is a distinctive position. It is illustrated by the science-fictional scenario where I q-proprioceive an *alien limb* with which I am not materially united, discussed in Chapter 4. There it was argued that on the materialist criterion, which says that a body-part belongs to a subject if and only if it is part of the material unity which they are, the limb is unquestionably yours. According to a self-conscious knowledge account, in contrast, a limb, materially united to someone else's body, but of which the subject has proprioceptive knowledge and can move basically, would not *clearly* be part of that other person's body. We saw that because the science-fictional case exhibits a dissonance between self-conscious knowledge and material unity conditions – they yield different answers to the question of whose limb it is – neither can one affirm that the limb is *clearly* part of the subject's own body.

This seems intuitive; if the other agent can move the hand basically, it is at least unclear whose hand it really is. (Though one could also ask how my movement of the hand could constitute a basic action.) One should therefore reject the materialist view that despite my having IEM-exhibiting knowledge of it, the q-proprioperceived limb clearly belongs to a distinct subject – while adding that that intuition does not imply that the limb clearly belongs to me.

This unclarity or indeterminacy is a result of the dissonance between self-conscious knowledge and material unity – the fact that the q-proprioceptive scenario seems to produce divergent answers to the question of ownership. If I have IEM-exhibiting knowledge of the limb – where two distinct subjects "share" a hand, for instance, without any suggestion that their two mere bodies constituted one lived Body – that limb does not clearly belong to a distinct subject, nor does it clearly belong to me. There is *no deeper fact of the matter*. The result is a refinement of (2):

> (3) X is *clearly* a body-part of mine *if and only if* either (i) I have IEM-exhibiting knowledge of it, and can move it basically; or (ii) it is part of a body-part that satisfies conditions (i).[11]

[11] To avoid complication I refer to the normal case – those with locked-in syndrome still have body-parts, and to cover such "hard cases" one would specify "Subjects normally have IEM-exhibiting knowledge", and so forth.

This is not an ad hoc stratagem to dispose of a difficult objection. It is a principled move arising from the breakdown in concept-application that arises from the dissonance between material unity and self-conscious knowledge conditions. Whenever these conditions come apart, there is an indeterminacy in the application of concepts such as bodily identity and limb-ownership.

A similar breakdown occurs in the case of monothematic psychotic delusion. There is evidence that X believes that p – apparently without intending to deceive, they repeatedly assert p – and evidence that X does not believe that p, namely, they act on the putative belief in circumscribed ways at best, and seem not to hold it on the basis of reasons. My thesis here is *not* that in many cases delusions are not beliefs, but rather that there is no fact of the matter concerning whether S believes that p.[12]

In strongly materialist accounts of bodily identity by Cassam and others, material unity is the exclusive criterion of bodily identity and limb-ownership; self-conscious knowledge is not a criterion. On the version of the self-conscious knowledge account advocated here, it is a criterion, though what it yields is normally congruent with material unity – and no clear fact about identity or ownership applies when it is not. One could perhaps say that materialism in principle rules out an essential indeterminacy in our talk of ownership of experience, one that arises from the possibility of conceptual breakdown. We will now consider why the purely materialist approach is inadequate.

5.2 Rejecting materialism concerning bodily identity and limb-ownership

The self-conscious knowledge account is now in its most plausible form, in which it should be preferred to materialism concerning bodily identity. *Materialism concerning bodily identity* regards subjects' bodies as essentially or fundamentally material objects, and criteria of bodily identity and limb-ownership as material. The materialist about bodily identity holds that a body-part is mine if and only if it is part of the material unity that I am. Despite its proponents' muted protests to the contrary, it follows that materialism in this sense cannot acknowledge that the

[12] "No fact of the matter" positions recur in this volume, and are presented in more depth in the discussion of delusion in Chapter 6. They were developed with regard to self-deception in Hamilton (2000), and with regard to delusion in Hamilton (2007a).

subject's immediate knowledge and capacity to move the limb basically are essential for limb-ownership. This fact marks a crucial divergence between materialism in this sense – and near-relations of that standpoint such as animalism – and the account defended here. Materialism fails properly to acknowledge that my body is individuated essentially as the body of a person, and not merely as a material or biological entity.[13]

An example of a materialist position is Cassam's. Cassam writes that his aim is to defend a materialist account of self-consciousness rather than a materialist account of the self. He defends what he calls an experiential version of materialism concerning self-consciousness, namely, that "we can and must be presented to ourselves, qua subjects ... as physical objects among physical objects".[14] This is the aforementioned *self-presentation thesis*, addressed from section 3 onwards. However, Cassam holds that the latter thesis may help support materialism concerning the self, the view that "the 'I' that thinks is ... shaped, located, and solid". If materialism concerning the self is correct, he writes, "then things are just as they seem in introspective self-awareness".[15] In this section I consider Cassam's sympathetic remarks concerning the metaphysical question of materialism concerning the self, and the associated positions of materialism concerning bodily identity, body-ownership, and limb-ownership.

I will focus on the question of limb-ownership. Here, Cassam initially seems to espouse an unqualified materialism that regards the IEM of judgements of bodily position and movement, such as "My left arm is hanging limply by my side", as at best de facto. That is, the subject can make judgements concerning their limbs that are mistaken owing to an error about *whose* limb it is. For instance, Cassam holds that in the science-fictional scenario where I receive information from an arm with which I am not materially united, my proprioceptive judgement will be mistaken owing to an error in identification. He later qualifies his position by attempting to acknowledge IEM, while still denying that "immediate presence" is as important as material unity for limb-ownership.[16] He now rejects the *tracking model* of self-identification whereby the subject genuinely identifies and tracks the position, posture, orientation, and

[13] Unless biology is humanistically conceived, perhaps; on which see Hamilton (forthcoming, 2013), Chapter 13.
[14] Cassam (1997a), p. 6.
[15] Cassam (1997a), p. 7.
[16] Cassam (1995), qualified in his (1997a). Bermúdez (2003) also regards proprioceptive IEM as de facto.

so forth of their own body, and the narrow sense of "physical object" which implies that model. In this way, Cassam believes, he has shown that materialism concerning the self – which at this point he labels "realism" – is consistent with IEM. He offers a further argument to reconcile materialism and IEM: if I am not materially united with a limb, it is only apparently "immediately present" to me, since "actual immediate presence requires material unity". So the judgement "My arm is raised", when based on the position of someone else's arm, must be the product of an illusion, and does not yield knowledge.[17]

The argument is too quick, however. It might appear that Cassam is wrong to suggest that the limb could "merely [seem] present to me", for "present to me" cannot be an externalist notion. But presumably what he means to say is something like: "Only apparently do I know on the basis of proprioception the position and so forth of the limb, for genuine knowledge requires material unity". What Cassam rules out is the possibility of proprioceptive knowledge concerning a material unity distinct from that of the subject. What my position rules out, in contrast, is proprioceptive knowledge concerning a distinct *person*. As the science-fictional example of the q-proprioperceived limb shows, these claims are not equivalent; where another agent can move basically a hand with which they are not materially united, it is at least unclear whose hand it really is. The materialist fails to acknowledge the intimate connection between bodily and personal identity which the q-proprioperceived limb scenario illustrates. To reiterate, they deny that there is a breakdown in concept-application where material unity and self-conscious knowledge conflict. For the materialist, the hand clearly belongs to the person with whom it is materially united. That, I believe, is mistaken.

In fact, the very statement of a criterion for limb-ownership is problematic for materialists. Cassam's definition that "for a limb or body-part to be a part of one, it is necessary and sufficient that one is materially united with it" invites the question what "one" refers to.[18] Clearly "one" is not meant to refer to "my torso", or "my mere body"; it must refer to a person. But it is strange to say "One is materially united with one's arm". *I* am not materially united with my arm; nor is my arm, strictly, part of me. If my arm is amputated, I have lost a part of my body, but – despite what Locke writes – it would be curious to say that I have lost a part of

[17] Cassam (1997a), p. 66. This is an application of Evans's argument against q-memory discussed in Chapter 2.
[18] Cassam (1997a), p. 66.

me. "Persons have parts" has no sense, neither in the case of temporal, nor of spatial, parts.[19] A better definition would be: "For a limb to be mine, it must be part of the material unity that I am". But I am not simply a material unity, and so my body-parts are not simply parts of a material unity. This a key problem for materialism.

Materialists face the further challenge, "What makes a given body my body?" They can hardly give the same response as they did concerning body-parts; it would be absurd to say that X's body belongs to X in virtue of being materially united with X. Materialism therefore seems to be in trouble. Cassam maintains that the challenge is illegitimate, since there is no body-independent "I" whose ownership of a particular body can be in question: "the 'me' can only properly be conceived of as a *bodily* me".[20] This may seem a good response, since it involves a salutary recognition by a materialist that – as Wittgenstein insisted – some philosophical claims or questions are not in good order as they stand, and require therapy for their correct elucidation. Certainly the claim that there is a "soul" to which one's body belongs, possessing criteria of identity independent of that body, is imponderable; Cassam is right to say that "the 'me' can only properly be conceived of as a *bodily* me". But "bodily" does not mean simply "materially unified". Although the challenge is not a demand for a criterion – "Which is my body?" is not something that ever needs to be decided – it does have a sense, namely one which invites the same response as the challenge concerning body-parts: "My body is the one concerning which I have IEM-exhibiting knowledge, and can move basically".

Cassam's rejection of the initial question "What makes a given body my body?" may be turned back on itself. The materialist reply assumed that the question made no sense because there was nothing to be decided. But neither – normally – does it need to be decided whether a given body-part is mine, yet materialists still offer criteria for ownership of body-parts. A gruesome exception where a decision is necessary would be where I am no longer materially united with the body-part in question, a limb for instance; other cases might be intertwined fingers, or a photograph of arms – one of them mine – with everything else blanked out.[21] But comparable photographs of torsos and arms would elicit the question "Which one is me?" rather than "Which one is my

[19] A claim defended in Lowe (1996), pp. 39–40.
[20] Cassam (1997a), p. 67.
[21] A more bizarre example is Oliver Sacks' "The Man Who Fell Out of Bed", in his (1985).

body?" Again materialists make an ascription to a body when it should be to a person.

Some varieties of materialism – though not Cassam's – might hold that the "one" with which body-parts have to be materially united is a brain. They fail to acknowledge that the brain is functionally defined as the creature's central control system; if there is nothing for it to control, then it is a mere mass of tissue, not a brain. In fact, probably the best answer that materialists can give is that "My body is the one by which I am materially constituted" – an answer incompatible with the account defended here. It is inadequate because reference to self-conscious knowledge is essential.

It is interesting that Wittgenstein contrasts the absurdity of claiming that there is a "soul" to which one's body belongs, with claiming that my nose is the one that is possessed by my body, which he does regard as in good order.[22] For although the latter claim does not seem absurd, it is questionable. As argued earlier, it is not "my body" which possesses a nose – rather, I possess it. It is part of my body, but possessed by me. Possession in the latter sense, in central cases, involves IEM-exhibiting knowledge and often basic movement, as well as material unification. (Often but not always basic movement, because as noted earlier there are many body-parts of which I have proprioceptive knowledge, but which I cannot move basically.) The centrality of cases involving proprioceptive knowledge and basic movement is what sustains the qualified self-conscious knowledge account.

It might be objected that materialism *is* compatible with the first key claim of this chapter, that "my body" is that of which I have self-conscious knowledge and can move basically. On this view, the objection continues, it appears that materialists can acknowledge the existence of persons, and so the required gap between materialism and non-materialism about the body does not exist. I would contest this line of argument, however. Insofar as proponents of materialism about the body acknowledge the role of self-conscious knowledge, to that extent they move away from materialism. There have indeed been strong materialists about the body who deny the importance of that knowledge; as we saw, this is a position which both Cassam and Bermúdez have espoused.

---

[22] Wittgenstein (1979), p. 24: "What is the criterion for 'This is *my body*'? There is a criterion for 'This is my nose': the nose would be possessed by the body to which it is attached".

Recent developments in medical technology may seem to challenge the self-conscious knowledge account. A "bionic arm" has been fitted to amputee patients. It is densely packed with microchips and miniature motors; microsensors pick up electrical impulses sent by the brain to absent arm muscles. The arm yields an analogy of proprioceptive knowledge and basic movement, and one patient says that he has "trained [himself] to think of combinations of movements to control [it]".[23] From the literature, it is not clear how far the subject has IEM-exhibiting knowledge of the arm, and can move it basically. But one can present the implications for ownership in either case. If the subject can move the arm basically, it is genuinely theirs; if they cannot, it is prosthetic, because it is not part of a body-part of which the subject has immediate proprioceptive knowledge. The inorganic material of which the arm is constituted might genuinely become part of the agent, if it became proprioceptively integrated in this way.

It might seem that cases where the subject lacks proprioceptive knowledge – either in part, due to local anaesthesia and proprioceptive deficit, or completely, in scenarios such as Anscombe's Sensory Deprivation Tank or locked-in syndrome – are more problematic.[24] However, the present account does not imply that in such cases the boundaries of my body fluctuate with present or actual bodily awareness. If I lose the ability to move it, the hand does not cease to be mine, since it remains part of a body-part of which I do have immediate knowledge and can move basically. To cover such "hard cases" as locked-in syndrome, one needs to specify that "Subjects normally have IEM-exhibiting knowledge", and offer an Aristotelian teleological defence. That large issue must be discussed elsewhere, and we now turn to materialism concerning self-consciousness.

5.3. Self-presentation versus elusiveness

In *Self and World*, as he explains, Cassam attempts to express some of the insights of Husserl and Merleau-Ponty within a Strawsonian framework. In so doing, he defends a version of materialism distinct from materialism about the self – what he calls the experiential version of materialism about self-consciousness, namely, "we can and must be

[23] Uhlig (1998). Clark (2003) discusses performance artist Stelarc's technological hand.

[24] Anscombe (1981b).

presented to ourselves, *qua* subjects…as physical objects among physical objects".[25] This I term the *self-presentation thesis*. Presentation, for Cassam, is an experiencing and not a conceptual relation. The more familiar version of materialism about self-consciousness, advocated by Strawson, Evans, and McDowell is a conceptual one. It says that I must conceive of myself, as subject, as an object (the always redundant "in the world" is usually added). Cassam argues that this position faces serious difficulties arising from the problem of misconception, and develops his own contrasting position in order to overcome it.[26]

Cassam deals painstakingly with objections to the claim that experience of embodiment is necessary for self-consciousness, but neglects the most fundamental criticism: of the very intelligibility of such "experience of embodiment". My aim is not to question the requirement that we experience ourselves as bodily entities – though it is difficult to see how an experience could itself be a permanent condition of experience – but rather the intelligibility of bodily self-presentation. I will argue that the self-presentation thesis constitutes a metaphysical inflation of the truism that subjects have immediate, self-conscious bodily knowledge which is IEM; it may be what Wittgenstein would describe as a misfiring attempt to say the unsayable. "I am embodied" is not a proposition based on evidence, but – at best – a grammatical proposition in Wittgenstein's sense.[27]

The thesis of self-presentation is not as easy to understand as may at first appear. Each of its components calls for analysis: I am presented to myself *as subject* (component 1); *as a physical object* (component 2). On Cassam's account, the first component, self-presentation as subject, involves various kinds of awareness: (1) awareness of oneself as a bearer of sensations; (2) awareness of that about which spatial relations are experienced as orientated;[28] and (3) awareness that does not require identification of a presented subject as oneself, that is, awareness on the basis of which it is possible to make first-person IEM-judgements.

---

[25] Cassam (1997a), p. 6. Together with rejection of the Exclusion thesis that the self is not an object among others in the world, this is his central thesis. Other proponents of self-presentation include Brewer (1995); O'Shaughnessy (1980); Husserl (1989); and Ayers (1991, vol. II, p. 286).

[26] Cassam (1997a), p. 9.

[27] This view is defended further in Hamilton (2013), Chapter X.

[28] "The physical entity that is the zero point of spatial perception presents itself as being a point of occupancy for psychological properties" (Cassam 1997, p. 57).

Cassam also mentions (4) awareness of oneself as a "subject-object" in Merleau-Ponty's sense, which he believes is equivalent to awareness of oneself as a person as defined by Strawson.[29] What these varieties of awareness imply is that self-presentation as subject is not just presentation of what is in fact the subject of one's thoughts and perceptions. Examples of such self-presentation as object rather than subject include touching myself, seeing myself in a mirror, and looking down at my body with that special restricted view engagingly depicted in Mach's *The Analysis of Sensations*.[30] In general, judgements made on the basis of visual perception are not IEM, and so do not fall under the heading of self-presentation as subject.

One could question whether proprioception is the only way that I experience myself as subject, as a physical object. When I remember climbing down Ben Nevis, I experience myself as subject, as physical object, not just as a mental object; I remember my legs aching. However, I wish to question whether one experiences oneself (as subject) as a physical object at all. To experience oneself as a physical object is, Cassam claims, to experience oneself as shaped, solid, and located; as satisfying Lockean criteria for physical objecthood. It is the bodily self or materially united whole which is "the *ultimate* presented subject of perception". It is the bodily self, rather than individual body-parts, that constitutes the presented subject; different forms of spatial perception are integrated with each other because their "local" points of origin are materially united.[31] The claim that it is the materially united whole which is ultimately presented is significant, as we will see.

5.4 Descartes is not committed to self-presentation

According to the dialectic that Cassam presents, the self-presentation thesis is opposed by the Humean *elusiveness thesis*, which denies that self-consciousness is awareness of the subject of one's representations as object. As Shoemaker puts it, "when one is introspectively aware of one's thoughts, feelings, beliefs and desires, one is not presented to oneself as a flesh and blood person, and one does not seem to be presented to one as an *object* at all".[32]

---

[29] Cassam (1997a), pp. 60–1, 77.
[30] Mach (1897), p. 16.
[31] Cassam (1997a), p. 66.
[32] Shoemaker, in Shoemaker and Swinburne (1984), p. 102.

I wish to undermine the opposition between self-presentation and elusiveness. This task is clarified by looking at the historical origins of the debate. We return to Cassam's contrast between materialism about self-consciousness, and materialism about the self, discussed in connection with limb-ownership. Cassam opposes *materialism about the self* with *the Exclusion Thesis* that the self is not an object in the world – an opposition at the ontological rather than experiential level, he believes. He argues that Descartes's account exemplifies the two independent levels of debate. The virtue of Cassam's interpretation is that it allows for a more sophisticated picture of Descartes's view of self-consciousness than the traditional one. Indeed, Descartes is a model for Cassam in this respect, except of course that while he is allegedly a materialist about self-consciousness, he is an *immaterialist* about the self. But we will see that Cassam's interpretation is only partly correct. Descartes maintains *immaterialism about the self*, on the basis of reason, but not materialism concerning self-consciousness.

Cassam recognises that for Descartes there is a deep tension between the deliverances of ordinary experience – on the basis on which I come to believe that I am an embodied creature – and reason, by which I know that I am in essence a thinking thing distinct from my body. Cassam describes Descartes's position as combining materialism about self-consciousness and immaterialism about the self. The tension is illustrated by a letter from Descartes to Princess Elizabeth, where he contrasts "the arguments proving the distinction between the soul and the body" with "the notion of the union which everyone invariably experiences in himself without philosophising", and continues: "Everyone feels that he is a single person with both body and thought so related by nature that the thought can move the body and feel the things which happen to it".[33] Thus Descartes maintains that although, according to experience, mind and body do not seem distinct, this perception is confused, so that they are conceptually and therefore really distinct. (According to some interpreters, he gives a scientific, mechanistic account of the interaction of psychic and physical events, and a subtler explanation of how mind and body coexist in one whole subject.)

We now see the error in the modern textbook view that, for Descartes, one intuits or experiences an immaterial subject. For in fact he holds that the immaterial subject is known by reason, not experience. This

[33] Descartes (1991), pp. 226–9. The tension is discussed in A. O. Rorty (1992) and MacDonald (2003), Chapter 7.

may not have been recognised by Hume, who in a famous passage assumes that Cartesians intuit an immaterial self:

> There are some philosophers who imagine we are every moment intimately conscious of what we call our SELF; that we feel its existence and its continuance in existence; and are certain, beyond the evidence of a demonstration, both of its perfect identity and simplicity ... To attempt a farther proof of this were to weaken its evidence; since no proof can be derived from any fact, of which we are so intimately conscious; nor is there any thing, of which we can be certain, if we doubt of this.[34]

The passage is not totally clear – indeed, "intuit" is perhaps ambiguous – but something of which "we are every moment intimately conscious" seems to be an object of experience rather than reason. So Hume seems to be guilty of what has become the textbook error, which assumes that Descartes intuits an immaterial self. So the latter's position becomes simplified.

This misinterpretation seems to have instigated a wild-goose chase into which even Kant may have been drawn. Strawson's succinct interpretation of the diagnosis presented in the Paralogisms says that the Cartesian mistakes the unity of experience for experience of a unity. If Strawson's interpretation is correct, then Kant regards Descartes as claiming that the subject has *experience* of an immaterial unity. This is a mistake, for Descartes claims only that the subject has an *idea* of such a unity.[35]

Merleau-Ponty, anticipating recent Descartes scholarship, tried to expose Hume's wild-goose chase:

> ... experience of one's own body runs counter to the reflective procedure which detaches subject and object from each other, and which gives us only the thought about the body, or the body as an idea, and not the experience of the body or the body in reality. Descartes was

[34] Hume, *Treatise*, I.iv.6. I am indebted to Peter Kail for the suggestion that Hume more likely has Malebranche, an unorthodox Cartesian, in mind. The *Treatise* is littered with paraphrases of Malebranche, who claimed, against Descartes, that we know the self by "inner sentiment" or "consciousness", since we cannot have a priori knowledge of the modes of substance.

[35] Strawson perhaps misinterprets Kant, who targets rationalist *arguments* concerning the soul.

> well aware of this ... [drawing] the distinction between the body as it is conceived through use in living and the body as it is conceived by the understanding.

However, Merleau-Ponty adds, for Descartes the body as conceived through use in living "remains subordinated to our knowledge of [the body] through the medium of ideas".[36]

A common criticism of Hume and to a lesser extent Kant is that they failed to consider the possibility of *bodily* self-presentation. It may seem that on the reading of Descartes that I endorse – namely, that he does *not* say that I experience an immaterial self – commentators have lamentably failed to recognise a possibility that he himself describes. Hence what has become a standard diagnosis of the error in Hume's account of the self, a diagnosis which – arguably – Cassam's treatment develops.

However, this is not what proponents of the revisionary reading should say. For Descartes did not really recognise the possibility of self-presentation. It is true that he claimed a tension between the deliverances of reason and experience concerning the self and self-consciousness, but rather than advocating materialism concerning self-consciousness, he maintains simply that I come to know or believe that I have a body on the basis of experience. This thesis does not amount to self-presentation, but constitutes a claim about the evidential basis for my belief or knowledge that I have a body.

As we will see later, even this more modest claim is misguided; I will argue that "I have a body", or "I am embodied", if intelligible at all, has no grounds. Thus the dispute over whether its grounds are furnished by experience or by reason is deflated. For immediate purposes, however, we can say that although Merleau-Ponty and Cassam write that Descartes does not propose an intuition or experience of an immaterial self, Cassam is mistaken in attributing to him a commitment to self-presentation.

5.5 The contrast argument against self-presentation

I have argued that the self-presentation thesis should not be attributed to Descartes. In fact, as I will now show, the thesis is faulty in a radical way; it is incoherent. The opposition between self-presentation

[36] Merleau-Ponty (1962), pp. 198–9.

and elusiveness constitutes a pseudo-problem, and so rejection of self-presentation does not lead ineluctably to the elusiveness thesis. The alleged dichotomy should be undermined: the assumption allegedly shared by Descartes and Hume, that it makes sense to ask how, if at all, I am presented to myself as subject, should be rejected as incoherent. The assumption is an expression of what Wittgenstein called grammatical illusion concerning the self, illusion arising from the truism that I have experiences of my body in a way that I cannot have experiences of someone else's body: I experience it "as subject". The self-presentation thesis is an unacceptable metaphysical inflation of this truism.

This is not to deny a modest and plausible line of thought, one that constitutes the limited truth in the self-presentation thesis. It refers to IEM-exhibiting judgement, cited by Cassam as a criterion for awareness of oneself as subject. The line of thought is this: proprioceptive judgements are IEM, and IEM is integral to self-consciousness. Non-proprioceptive knowledge of one's body comprises ways of knowing that are not IEM-exhibiting, and so not ways of knowing oneself as subject – or as I would prefer to put it, they are not distinctively self-conscious ways of knowing. If proponents of self-presentation limited themselves to this line of thought, their anti-immaterialist and anti-Humean conclusion would be plausible; certainly it yields a valid critique of Hume's account. This is not to concede that what is missing in Hume's treatment is recognition of bodily self-presentation; it is simply to say that there can be self-conscious knowledge of one's own body. However, proponents of self-presentation are not just claiming that there is self-conscious knowledge of one's own body. They wish to make a more expansive claim about experience of the "materially-united whole" that they see contemporary proponents of elusiveness as denying; a more modest claim could hardly be opposed by the elusiveness thesis which Cassam rejects.

The self-presentation thesis is shown to be objectionable by means of a *contrast argument*, according to which there is no intelligible contrast to experience of self-presentation – an objection that Cassam does not address directly. One might question whether there has to be a contrast; the way experiences are might be the way they have to be. But this rather Kantian formulation is not how Phenomenologists usually put the claim – nor how Cassam puts it. He does suggest three contrasting "grades" of "apparent presence" in the world: (a) intuitive awareness of oneself as present in the world only geometrically, as something bodiless; (b) awareness of oneself as a point of view which

"has" a body which plays a special role in determining the course of one's experience, but which does not amount to being embodied, that is, (c); (c) awareness of oneself qua subject as a bodily presence in the world, that is self-presentation in Cassam's sense.[37] Do (a) and (b) offer the required contrast with (c)? Grade (a) seems impossibly abstract and imponderable. What could determine a subject's location if not their body? Clearly one cannot assume that, without a body, one has eyes and ears from which to view. Imagining looking at a room from a merely geometric viewpoint is the same as simply imagining the room.

Grade (b) expresses a Cartesian mechanistic standpoint derived from the traditional textbook interpretation of Descartes, and it is useful to pursue this possibility a little. One of its clearest advocates was Thomas Reid, who subscribed more faithfully to the textbook Cartesian treatment of the body than did Descartes himself. On Reid's view, I really am lodged in my body like a pilot in a vessel, as Descartes put it in the *Meditations*; for Reid, the body is the instrument of the mind.[38] Descartes himself, as we have seen, was more equivocal, maintaining that although experience shows that mind and body are not *really* distinct, they are nonetheless *conceptually* distinct. I will argue that candidates for (b) – notably, subjects with severe proprioceptive deficit – in fact exhibit an attenuated form of (c). It follows that the idea that there is experience of (c) is not intelligible, since there is no genuine contrast with a corresponding lack of experience of (c).

The required contrast, which I believe is not available, is experience that is not bodily in an IEM-exhibiting way – where I do *not* experience myself as subject, as bodily object. That there must be such a contrast follows from the status of experience as something of which one is conscious – ascription of an experience implies the possibility of it beginning, ceasing, or not occurring at all. A conception or idea may be tacit, but an experience cannot be. This is shown by the example of persistent tinnitus. The subject begins to notice noises in the ear when formerly they had not. Some medical explanations hold that everyone has inner noises, but only tinnitus sufferers actually notice and become concerned by them; however, even if one thinks of the experience as being present from birth, one can at least imagine it ceasing. This is

[37] Cassam (1997a), p. 58.

[38] "…the body, as a part of the man, and an instrument of the mind…" (Reid (2002), p. 613, Essay VIII, "Of Taste").

not so with the alleged experience of being embodied – one could not imagine such an "experience" ceasing until the end of all experience in death, I will argue.[39]

Another way of formulating the issue is through a popular question in recent philosophy of mind – Thomas Nagel's "What is it like to be … ?" question. The self-presentation thesis assumes that there is something it is like to be embodied. Now this question needs to be treated sceptically, I believe, as Norman Malcolm did when he asked, half facetiously, whether Nagel's question "What is it like to be a bat?" is like the question "What is it like to be a lorry driver?" That is a perfectly sensible question, to which an answer might be: "It's a hard life. You get up early in the morning and work long hours. It's monotonous work, stopping off at dreary motorway cafes, it's badly paid, etc." Non–lorry drivers can learn something from this answer. "What it is like" is not, in this case, something ineffable, as Nagel mistakenly supposed it was in the case of "being a bat" or "being a human".[40] Can one give a similar response to the question "What is it like to be embodied?" Consider the answer: "I feel the pressure of my buttocks on the seat I'm sitting on, I know immediately the position and location of my arm, I get backache … is that what you mean?" No one hearing this reply – even someone with proprioceptive deficit – will be unaware of these sensations or others like them, so no one will learn anything from their description. I have particular experiences of my body which begin and cease, and it is these which the answer to Nagel's question cites. I place my hand on my leg and feel its touch and pressure; I withdraw my hand and cease to feel it. But it does not follow that I have "experience of being embodied", since that is not something which I could cease to have. Particular experiences of my body do not add up to experience of being embodied, experience of the materially united whole.

The putative experience of the materially united whole cannot simply be the sum of all the particular experiences of one's body, some of which patients with proprioceptive deficit lack. That such a set of

---

[39] Contrast arguments have been employed, for instance, in discussions of the problem of evil. It could be argued that the existence of evil does not imply the existence of good, but that the possibility of experiencing or conceiving of evil does imply the possibility of experiencing or conceiving of good.

[40] Malcolm's comments were made during a seminar series at Kings College London, 1981–1982. Nagel's discussion is in Nagel (1979). Hacker (2002) provides a persuasive critique of his views.

experiences exists could hardly be what the Hume-Shoemaker elusiveness thesis denies; rather, it denies that I experience myself bodily *as subject*. The rival theses purportedly concern experience of the whole self, not particular experiences of body-parts. It follows that as suggested earlier, "I am presented to myself qua subject as a bodily entity" can only mean: I have immediate, IEM-exhibiting knowledge of my body and its parts. The truth of this statement – which is surely a truism – is obviously not what proponents of self-presentation or elusiveness think that their debate is about, otherwise Shoemaker's claim that "one is not presented to oneself [as subject] as a flesh and blood person" would be absurd. Maybe the claim is absurd anyway – and certainly Hume's famous claim that I cannot introspect a self may be countered by the truth that I have IEM-exhibiting self-conscious knowledge of my body through proprioception. Yet there is no intelligible alternative to the truistic interpretation.

It may be argued that there *are* possible contrasts to the alleged experience of being embodied.[41] But out-of-body experiences deserve an A. J. Ayer–like scepticism, of the sort Ayer himself displayed until apparently he had such an experience himself and reported it in one of the Sunday newspapers.[42] Proprioceptive deficit offers a better prospect of a contrast, and indeed one partially subject with proprioceptive deficit has urged that there is an "experience of being embodied" which he has in various respects lost.[43] If there were a condition of complete, congenital proprioceptive deficit, perhaps sufferers might learn something from the description of what it is like to be embodied. Oliver Sacks' case study of Christina offers the closest to a complete contrast that proponents of self-presentation could claim. Her deficit is so radical that she describes herself as feeling "disembodied", and says that she uses her body as a "tool".[44]

Even in such severe cases, however, there is a proprioceptive residue of sensations of pain, fatigue, and temperature; limited proprioception of the upper body; and a "point of view". Christina does not use her entire body as a tool – no one could – but just her torso. There is a proprioperceptive residue even for victims of "locked-in" syndrome, who have suffered a devastating stroke, and comatose patients whose consciousness goes unrecognised because they are incapable of action.

---

[41] This is what Cassam urged when I suggested this line of objection.
[42] A curious episode discussed in Cash (2001).
[43] Personal communication from Ciarán Benson.
[44] In Sacks (1985).

Action would be impossible without this residue; Ian Waterman could not have learned to move his body through visual perception only, without being able to move his head and eyes. Indeed, as was argued at the end of the previous chapter, visual perception requires proprioception.

This is not the sense of "experience of being embodied" advocated by proponents of self-presentation, therefore. As well as the fact of a proprioceptive residue, note also that Christina describes herself as feeling "disembodied" – that is, formerly embodied – not "non-embodied". Though it sounds odd to say it, this lack of experience does have an experiential quality. So the question is whether it is possible for a subject to have total, congenital proprioceptive deficit. Consider again Cassam's grade (b) of presence in the world: awareness of oneself as a point of view which "has" a body, a body which plays a special role in determining the course of one's experience, but of which one's experience does not amount to awareness of oneself qua subject as a bodily presence in the world. For (b) to yield the contrasting experience – namely, lack of experience of embodiment – it must be had by a self-conscious subject, and it is doubtful that the (b)-subject is self-conscious, or even conscious. Indeed it is doubtful that they have experience at all. So proponents of the self-presentation thesis face a dilemma. Either the experience of being embodied is not completely eradicated, in which case the contrast has not been provided, or it has been, but there is a question whether the resulting subject is self-conscious or even conscious at all. Clearly the (b)-subject – a subject that has experiences but lacks experience of embodiment – must be more than a robot or machine.

Materialism concerning self-consciousness is therefore more properly expressed as the view that experience of embodiment is necessary for self-consciousness. But on this view, it looks as if proprioception is not merely a form of self-conscious knowledge, but is in fact necessary for perception and therefore self-consciousness – from which it would follow that one cannot conceive of an experiencing subject which lacks the experience of embodiment.[45] One reason why a creature which lacked proprioception could not perceive is the vital role of visual proprioception discussed in the previous chapter. Visual proprioception, Gibson argued, enables the subject to differentiate between a

[45] In a more recent article (Cassam 2002), Cassam does argue that awareness of one's own body is a necessary condition of acquisition and exercise of concepts of primary qualities of objects.

change of place and a change of state of an object – between changes in perceptual experience that are reversible by movement of the observer (back to the original position of observation) and those which are not. Without proprioception, a creature could process visual information, if at all, only so crudely that one could not really speak of perception and consciousness. It could not move actively, and would be unable to distinguish the results of passive movement through a stable environment from genuine change in that environment. Seeing is an active faculty that involves doing something with the visual information, and thus involves action and proprioception. The ability to move the eyes is not essential, but moving one's head or even whole body would be. A static creature with an "eye" – a lens, that is – could not, therefore, be conscious.

Cassam's position is encouraged by thinking of the complete lack of proprioceptive knowledge as comparable to the complete lack of visual knowledge experienced by totally and congenitally blind subjects. This approach treats proprioception as a sixth sense. But as preceding arguments have suggested, lack of proprioception is a far more radical deficiency than blindness. This fact supports the view defended in the previous chapter, that proprioception is not one of the senses, but is a non-sensory, primordial capacity. Subjects with proprioceptive deficit report the lack of experience of embodiment on the left side. But proprioception is about feeling my arm, leg, and so forth, not about feeling embodied. Cassam assumes that one can extrapolate from the experience of partial lack of bodily feeling to experience of complete absence of feeling of embodiment. But although one can make sense of a partial lack of the experience of embodiment, one cannot make sense of a complete lack – recall Ryle's remark that not all coins could be counterfeit.

One corollary of the self-presentation thesis is the claim that one experiences oneself as solid. The response to this claim provides further illustration of the contrast argument. Cassam argues that solidity is typically felt as an impediment to one's movements, "and to experience a solid object as an 'obstructive something' ... is at the same time to be sensibly or intuitively aware of that which is obstructed – the *subject* of tactile perception – as something solid".[46] Now certainly I have no thought that I might not be solid, hence my lack of surprise when my hand does not simply pass through or envelop solid objects,

[46] Cassam (1997a) , p. 52.

but exerts a force on them and often moves them.[47] Subjects do not normally, after early childhood, attempt to squeeze through impossible apertures or stand on objects that have no chance of supporting their weight. These facts perhaps show that I conceive of myself as a solid object of a certain shape and size. But they do not imply that I experience my hand as solid. This is because solidity is typically assessed *by* my hand or other parts of my body, the solidity of which is assumed otherwise they would be useless for the purpose. These body-parts are not themselves experienced as solid, except in special reflexive cases – precisely those cases where I treat myself as object rather than subject. For instance, when I press my hand against my thigh, I may certainly experience my hand or thigh as solid, but I have to attend to one or other of these "double sensations" in order to do so. To move my hand basically, is not to treat it as an object. I feel objects as solid because there is a contrast; I am used to moving my hand through thin air, for instance. This is not so with regard to feeling my hand as solid.[48] There is a contrast with *feeling other things* as non-solid, but not with feeling them *with* something non-solid, since I could not have the experience of being immaterial or ghostly.

If the required contrast were available, then the self-presentation thesis would indeed amount to the claim that proprioception is a form of self-conscious knowledge. My conclusion, however, is that it is very doubtful that there is experience of being embodied in the sense which that thesis requires.

Here ends the direct critique of materialism concerning self-consciousness, though the next section argues against it indirectly.[49] My conclusion is that the self-presentation thesis is a confused development of the truistic claim that conceiving of oneself as a person among persons is integral to self-consciousness. Note, however, the distinction between

---

[47] Of course a solid hand might pass through a ghost, or vice versa. But Judge Peter Stretton's robust attitude to a case involving a haunted house is commendable: "Ronald Sowter … claimed that during a visit … he had felt as though 'something walked through' him. The judge pondered: 'One poses the question – what does it feel like having someone walk through you? The answer to that is that nobody knows. How can he not say it was not simply a draught of air or a sensation from his own body rather than a ghost?'" ("The ghostbuster judge", *Daily Telegraph*, 19 January 1999.)

[48] Similar arguments to those concerning solidity are applicable to location, discussed by Cassam (1997a), pp. 52–3.

[49] Chapter 6 attempts to undermine Cassam's problem of misconception from a different direction.

conceiving of oneself as a person – the Strawsonian description – and conceiving of oneself as embodied. "Conceiving of myself as a person" is a much broader notion than "conceiving of myself as embodied", and is a more acceptable one, I would argue. It follows that Strawson's original account is to be preferred to Cassam's development of it. What truth there is in materialism concerning self-consciousness – though to reiterate, the use of the term "materialism" is not satisfactory – is that self-conscious subjects must conceive of themselves in this way. Even this claim should be treated with caution, however, as we see in Chapter 6 – which argues that the subject's self-conscious self-conception is essentially negative.

5.6 "I know that I have a body" and Anscombian doubt

The self-presentation thesis, and the claim that I know that I have a body, are philosophical inflations of the truism that as subject, I have proprioceptive knowledge of my body. As discussed earlier, Descartes claimed that, through ordinary inner experience, we come to believe that we are embodied creatures, while our reason tells us that we are in essence thinking things. I will argue, against Descartes, that "I have a body" is not subject to confirmation, and indeed is not an item of knowledge, but is a piece of latent nonsense, as suggested by Wittgenstein in *On Certainty*.[50]

Surprisingly – given that she edited *On Certainty* – Anscombe in her article on the first person does not consider this position. What I will call *Anscombian doubt* rests on the Cartesian assumption – implicit in her discussion of "I" – that I know that I have a body on the basis of experience. Her Sensory Deprivation Tank Argument, which concludes that "I" is not a genuinely referring expression, envisages a subject in a state of complete sensory deprivation and locally anaesthetised all over, who as a result doubts whether they have a body.[51] The plausibility of the Anscombian doubt is undermined by principle (E):[52]

(1) (E) It is not merely by current sensory experience that I know that I have a body (i.e. past experience suffices).

---

[50] Wittgenstein (1969b).

[51] Short-term relative sensory deprivation in a "flotation tank" is, in contrast, benign, and is used as relaxation therapy (Brown 1989).

[52] This response is based on a suggestion by Kenny (1979), p. 12. Anscombe's Tank Argument is subjected to closer criticism in Hamilton (1991).

So (2) Lack of present sensory experience in the Tank cannot ground a doubt about whether I have a body.

Alternatively, the subject may be portrayed as thinking "I doubt whether I still have a body". Either way, Anscombe's argument requires the further condition that her victim is amnesiac; there is no recollected sensory experience either.[53] The postulation of amnesia in addition to total sensory deprivation means that Anscombe's opponent has to rely on the more contentious principle (E*):

(1) (E*) It is not by sensory experience that I know that I have a body.

So (2) Lack of sensory experience in the Tank cannot ground a doubt about whether I have a body.

Some may consider that principle (E*) is obviously false; if not by sensory experience, how else do I know that I have a body? But although sensory experience may be a cause, it is not a ground of my belief that I have a body. (E*) asserts that it is not the role of my experience to confirm or provide evidence for the fact that I have a body. It should be noted that (E*) is not a consequence simply of the enormous evidential over-determination of "I have a body". Sensory experience could not, in the normal course of events, be said to confirm that I have two legs. But lack of any sensory input from one's leg, for instance in a limited version of the Sensory Deprivation Tank scenario, might disconfirm this. Indeed any unsurprising observation may have negligible or non-existent confirmatory value – "This raven is black" hardly confirms "All ravens are black", in the standard example – while its perhaps startling negation may have high disconfirmatory value. Such cases exhibit an asymmetry between confirmation and disconfirmation. That is, the inference

If not (p confirms q)

Then not (not-p disconfirms q)

does not hold. Furthermore, following an extraordinary apparent disconfirmation, it will make sense to talk of (re-)confirmation. Thus one may argue that a doubt that I have two legs, entertained while in

[53] It is presumably this premature refutation that leads Rovane (1987) to talk of "Anscombe's amnesiac under sensory deprivation".

a state of sensory deprivation, is dispelled by the discovery that my leg has been locally anaesthetised. Might the Tank-victim's belief "I have a body" similarly be re-confirmed by the return of sensation to their totally locally anaesthetised body?

It is not simply evidential over-determination that rules out an initial confirmation in the latter case, however. The reason why "I have a body" is not confirmed by experience is that for creatures to whom one can ascribe the concept "my body", acquisition of sensory experience necessarily accompanies a developing conception of oneself as an embodied subject. One could not acquire the concepts exercised in sensory experience without at the same time acquiring the concept "my body", and vice versa. This is a further example of the conceptual holism central to this volume. Since the very having of sensory experience presupposes that I have a body, it cannot provide evidence for this fact. This is not to say that having sensory experience must involve the exercise of concepts; in the case of infants and animals, it may not. But the belief that I have a body – if it is a belief, rather than a grammatical proposition in Wittgenstein's sense – clearly does involve the exercise of concepts, notably "my body"; any experience that could provide evidence for the belief would involve concepts also. My belief that I have a body is not grounded in sensory experience, nor in anything else.[54] One should conclude, therefore, that although deprivation of sensory experience may have a bizarre effect on the subject's network of beliefs, such a deficit merely causes and does not justify the Anscombian doubt. (Which in any case could equally be "Am I alive?")

These considerations are unlikely to persuade those who, wittingly or not, adhere to a Cartesian conception of mind, however. This category turns out to include Anscombe herself, since as noted earlier the Anscombian doubt is essentially Cartesian. This becomes clear when one realises that the subject is justified only in thinking the following:

(i) "Nothing in my present or recollected experience is experience of a body"

[54] The status of "I have a body" is explored in Hamilton (forthcoming 2013), Chapter X. Cassam's view that sensory experience has a "bodily character" ought to make him sympathetic to the conceptual holist claim, but he seems not to be.

They are not justified in inferring from (i), the stronger claim required to motivate the Anscombian doubt, namely, either:

(ii) "This is what I would experience if I had no body"

Or perhaps:

(iii) "I have had, and am having, an experience as of lacking a body"

Thus the doubt, and the anti-Cartesian no-reference view of "I" which Anscombe believes may be inferred from it, are mutually undermining. The line of thought that gets us from (i) to (ii) or (iii) is this. The particular bodily experiences that I have, the aches and pains, feelings of warmth and coldness, of the wind blowing on my face, of moving my limbs, and so on, together make up the experience of being embodied: Cassam's notion of bodily self-presentation criticised earlier. So when I am deprived of all such experiences, the argument continues, I have experience as of not being embodied – experience as of lacking a body.

It may be urged that there is such a thing as the experience as of lacking a body, since it is precisely this that the Tank-victim, and subjects with severe proprioceptive deficit, are having. But for the reason just given – and those presented in the earlier critique of self-presentation – this is not so. The Tank-victim is an embodied subject who is experiencing an absence of current or recollected information from their body. This experience cannot be equated with that which a subject would allegedly have if *per impossibile* they lacked a body. One might say, tracing Anscombe's argument back to its Cartesian origins – though she evidently wishes to remove its sceptical associations – that Descartes only thought that he could imagine being disembodied. But "I have a body" and "I am embodied" are genuinely puzzling propositions. To be embodied is not merely to be related to a body, but should one say that I am identical with my body? The relation between oneself and one's body – if one can so much as talk of a relation – must be a primitive one.[55] "I have a body" is no more meaningful than "I am a body". An illustration with the

[55] There is much to be said on this topic. Baker for instance attempts to compare the relation of the person and the human animal that constitutes it, with that between a statue and the piece of clay of which it is composed, or a flag and the piece of cloth of which it is made (2000, Chapter 2), but the account I have been developing stresses that it makes a difference that unlike the statue, the person is an essentially self-conscious entity.

caption "I have a body like this", put into a space capsule and directed at inhabitants of distant galaxies, is intelligible; the caption "I have a body", with the implied contrast "unlike you who may not", is not.

5.7 The Body-body problem

As a coda to discussion of proprioception and the body, some mention must be made of Arthur Danto's intriguing suggestion of a *Body-body problem*, which seems to challenge the claim of a conceptual holism governing proprioception and the lived Body.[56] The Body-body problem concerns the relation between "my body" and the mere body – between bodily ascriptions to persons, and predications of human bodies which are not also ascriptions to persons. This problem seems to displace the mind-body problem because consideration of the lived Body brings into question the very definition of "mind".

To say that the body is a physical object is just to say that predicates may be applied to it that apply with just the same sense to non-persons. Thus there is a sense in which it has mass, chemical constitution, size, and traces a spatio-temporal path. "I weigh 12 stones", "I am 6 feet tall", and "My body contains X litres of H_2O" each involves predicates that apply with the same sense to non-persons, as, for instance, in "My bath contains X litres of H_2O". As Husserl puts it, the Body is "integrated into the causal nexus of material nature".[57] "Holding the block of ice made my hands numb" or "The weight of the stone pressed down on my leg" are examples. For physiologists, "physical body" seems to be the primary sense of "the body"; their interest is with the "mere" body.

It is not quite right to assume, with Merleau-Ponty, that the Body-body problem concerns the divide between the first and third persons. He does write:

> It is indeed not enough to say that the objective body belongs to the realm of "for others", and my phenomenal body to that of "for me", and we cannot refuse to pose the problem of their relations, since the "for me" and the "for others" co-exist in one and the same world. ...[58]

But Merleau-Ponty is wrong to talk, in the first instance, of a "for me" and "for others". One can make physical attributions of one's own

[56] Danto (1999d).
[57] Husserl (1989), p. 167.
[58] Merleau-Ponty (1962), p. 106.

body; one can treat it merely as a physical object, for instance, heroically, "I interposed my body between my partner and the gunman", or "I used my body heat to try to prevent the child suffering from hypothermia, by holding it close". This is to recognise and exploit the fact that my body has physical properties like other animate or inanimate objects: the ability to act as a barrier, possibly with fatal consequences, or as a heat producer. But it is not the case that one always does this when acting.

However, although it does not concern the interaction of distinct substances, the Body-body problem, like the mind-body problem, may be a pseudo-problem. Certainly, it would be a mistake to regard "Body" and "body" as two distinct entities. The position I will defend may be understood by considering the alternatives to it:

(1) A Cartesian mechanistic account, in the textbook sense. This simply resurrects the mind-body problem, so it is more a rejection than a resolution of the Body-body problem. A clear advocate is Thomas Reid, who, as we saw earlier, in some respects subscribes more closely to what is traditionally regarded as a Cartesian position than does Descartes himself. For Reid, I am lodged in my body like a pilot in a vessel, and the body is the instrument of the mind.[59]
(2) A Fregean "different sense, same reference" account, which holds that the lived Body and the mere body are the same object viewed in two different ways. This is suggested by Husserl, who writes that from a phenomenological point of view, "the physical and the lived Body are essentially different", though they are the same object. Merleau-Ponty says that the objective body and the phenomenal body are two aspects of one and the same thing. But he also says that the "body-subject" involves a part–whole relationship between physiological elements and the embodied subject; the whole is more than the sum of the parts, rather than one object with two aspects.
(3) The possibly materialist claim that there is one sense of "body", qualified by "living" and "dead"; the lived Body/mere body distinction simply expresses the fact that the body has certain capacities when the person is alive, which it loses when they die. It is "body" in the sense of "corpse" which is the primary sense in which the body is a merely physical entity.

[59] "...the body, as a part of the man, and an instrument of the mind..." (Reid (2002), p. 613, Essay VIII, "Of Taste").

The Cartesian mechanism of (1) is unviable; (2), though more plausible, is not correct either. Although it is right to say that I have proprioceptive knowledge, as lived Body my body is subject rather than object. On the Fregean view, "Morning Star" and "Evening Star" have different senses but the same referent because the methods of viewing are essentially similar: observation of Venus at different positions in the sky. The sense-reference distinction applies only to objects with similar identity and individuation criteria; stars (Morning and Evening), persons (Cicero and Tully), and mountains (Everest and Gaurisankar), but not "pain" and "C-fibre firings", or "Caesar" and "0". In the latter cases, no appropriate sortal concept covers them both. "Object", "event" and "process", in the very abstract philosophical sense, are not appropriate sortal concepts. One cannot talk of criteria of identity for anything as general as "object". The Morning Star and the Evening Star can be described as the same object, only because one can make the prior claim that they are the same heavenly body.

This leaves option (3). The body does have certain capacities when the person is alive, which it loses when they die. Also in favour of (3) is the fact that, although criteria for individuating senses are notoriously vague, "body" does not seem to be ambiguous except perhaps in the ordinary sense in which the (living) body is contrasted with the corpse. So the difference between the lived Body and the mere body is not that between two senses of "body". However, this option fails to recognise the use of talk of the "lived Body", "body-subject", or "bodily self", in showing the diversity and significance for self-consciousness of what are traditionally regarded as uniformly physical ascriptions to persons. Expressions like "lived Body" are reminders that many such ascriptions – those involving Strawson's "M-predicates" – are not strictly physical, but belong to the ambiguous category outlined here. They are, however, ascriptions to persons and not bodies.

The lived Body is the primary concept; the corpse is the ex-lived Body. As Merleau-Ponty argues, the mere or physiological body is a theoretical construct or abstraction, most apparent to the subject in cases of illness where capacities are restricted, or in physiological investigation. Such is the divide between living Body and corpse that the idea that criteria of bodily identity connect the two, and give equal treatment to each, sounds odd. The criteria of bodily identity are essentially dependent on those of personal identity; indeed, talk of "bodily identity" may not have a clear sense.

Talk of the "lived Body" is metaphorical, a way of emphasising certain kinds of ascription to persons, and perhaps in this sense the Body-body

problem is a pseudo-problem. "Lived Body" is taken over-literally in David Bell's presentation of Husserl's concept. It is, Bell claims, immediately expressive; is sensitive, to pain, heat and cold, and so on; has motility or power to act; and functions as the absolute point about which all spatial relations are experienced as orientated. These are all said to be properties of the lived Body; it is this which yawns, smiles or cries out, or has the power to act.

However, these are all things that *I* do – not my body, lived or not. Perhaps one can say that the mere (physical) body is sensitive to heat and cold – and so I can get sunburned or frostbitten – and in a different sense that I am. But it is not my body that is sensitive to pain; rather, I am. Attributions to the lived Body are, strictly, part of the range of attributions to persons. (Of course proponents of the lived Body do not deny that there are attributions to subjects which are not also attributions to the lived Body, for instance, thoughts.) There is no need to postulate another bodily entity or aspect of bodies, separate from that of persons.

One can respond similarly to Merleau-Ponty's concept of the body-subject. He implies that the proper distinction is not between different kinds of object – phenomenal and physical bodies – but between the body as object, and as subject: "I observe external objects with my body, I handle them, examine them, walk round them, but my body itself is a thing which I do not observe". Also: "My body is constantly perceived *[cf. observed]*. It is therefore an object which does not leave me. But in that case is it still an object?"[60] As a corrective to Cartesian mechanism this claim has some justification, but it remains the case that I, and not my body, am the subject. The term "body-subject" simply captures the fact that central varieties of bodily self-ascription are immediate and IEM.

5.8 The unity of consciousness versus the unitary self

We must finally relate discussion of proprioception and the body, with earlier discussion of the opposition between unity of consciousness and unitary self. To reiterate, the claim of a conceptual holism between memory, expectation, and personal identity is the denial – to use more grandiose terms – that the unitary self has precedence over the unity of consciousness. This is the unity of consciousness as a non-Humean necessary unity, not the unity of the stream of consciousness: the contiguity of one's experiences, simultaneous or diachronic, of a beautiful sunset, a

[60] Merleau-Ponty (1962), p. 91; see also Bell (1990), p. 208.

pang of hunger, and a moment of melancholy reflection. It is the unity of the remembering or anticipating subject defined in terms of IEM, that between the subject's recollection and the recollected experience, or the subject's expectation and the anticipated experience; it consists in the fact that either the subject o-ed or will o, or else no one o-ed or will o.

A non-Humean unity of consciousness standpoint says that what unifies a single set of experiences is simply that they are all experiences had by one person; there is no further, underlying fact of the kind assumed by proponents of the unitary self. Since the question, "What is it that unites a person's present experiences with their past ones?" presupposes that it is the same person, there is nothing further to say after one has answered the identity question "What are the criteria for personal identity?"[61] However, one should add that proponents of the unitary self, such as A. J. Ayer and Edward Craig, are wrong to maintain that the identity question is more fundamental than the unity question. They hold that experiences are united not simply in virtue of being had by the same person, but more fundamentally by standing in a certain causal relation to the body or brain by which the person is identified. Since Ayer takes "body" as more fundamental than "person", he answers the unity question by giving a particular answer to the identity question; there is a "further fact", a material one. He answers the question "In virtue of what are experiences had by the same person?", by claiming that "a person's ownership of states of consciousness consists in their standing in a special causal relation to the body by which he is identified".[62]

Their responses to the unity question therefore define the opposition between unitary self and unity of consciousness standpoints. Like Ayer, Craig assumes the unitary self standpoint, which says that one can first identify the body and then assign the experiences. In his discussion of Hume's account of personal identity, Craig asks:

> Why do I not then take some of your mental states for mine? ... The answer, obviously, is that your perceptions are no part of the input

---

[61] The terminology of "unity" and "identity questions" is found in Vesey (1977).

[62] Ayer (1963), p. 116. He sympathises with Strawson's claim that "person" is a fundamental category, but develops an eclectic position which also borrows from reductionism. The kinship which he claims with a Wittgensteinian "no-ownership" position is not obvious, however. His view may therefore be regarded as an unusual synthesis of materialism and idealism. At a different phase of his career, Ayer advocated a reductionist version of the unity of consciousness standpoint.

> to the device which produces my belief in the unity of my mind. It is not as if there were a mechanism which was allowed first to survey all the perceptions there are, and then parcelled them out ... [63]

The arguments of the present chapter show that this answer is not obvious at all. Of course there is no parcelling mechanism which surveys unowned perceptions. But the "device" which produces my belief in – or rather, assumption of – the unity of my subject, cannot be individuated independently of the subject which has the belief that it is unified. The device that Craig has in mind, presumably, is the brain. However, the brain in question has to be "my brain", and so there is an unacceptable circularity in his response. Although it undermines Craig's position, this circularity does not undermine the unity of consciousness standpoint. It is consistent with the Wittgensteinian interpretation of "Another person can't have my perceptions", as a grammatical proposition and not a metaphysical truth. This interpretation says that it is a matter of grammar that the criteria of individuation of perceptions make essential reference to their possessor.[64]

There is a further important consideration that Craig neglects. The "device" (the brain) which receives perceptual input is part of a unified functional system; the brain and nervous system should be defined functionally as the subject's bodily control system. According to the personal conception of information defended in Chapter 2, what counts as input, for a self-conscious subject, depends on their perceptual self-ascriptions; subjects cannot be individuated by brains and nervous systems construed in physicalist terms. Thus there is a conceptual connection between "brain" and "person" – or, more broadly, "intelligent living creature". A slime-mould aggregate could not have a brain. Saying, of something that looks and appears to act like a human being, that it has no brain, is tantamount to saying that it is not a person – that it is a robot with no autonomous control. Thus there is a neglected line of objection to sceptical brain-in-a-vat thought experiments, that they rest on divorcing the "brain" from its function as central control system; when "brain" is defined functionally, there *is* no brain in a vat, but simply an organic material unity. So it could not turn out that I am just a brain-in-a-vat, because whatever it is that is in the vat, it is no longer a brain. The sceptic's argument seems to be: "All these events seem to

---

[63] Craig (1987), p. 112.

[64] Wittgenstein's own discussion concerned pain, of course (Wittgenstein 1953, para. 253).

happen to you, but how do you know that they *really* happen, rather than just seem to happen?" To this, the proper response is: "If none of them really happened, there would be no *me* that they merely seemed to happen to".

Some common misconceptions about the unity of consciousness standpoint should be cleared up. It does not imply that the subject actively unifies itself, nor that it may sometimes be in doubt about whether it is a unified subject. It makes no sense for me to ask whether I am a unified subject – whether there is one object which I am, one living human being – and thus it could not be an item of knowledge for me that I am. The very concept of a subject of experience assumes unification. "Is there just one object with which I am identical?" is at least pragmatically self-defeating, since an affirmative answer is presupposed by a comprehending use of "I".[65] Consider Campbell's parallel between self-knowledge and a subject watching a set of closed-circuit TV screens. There is no need for the viewer to use internal evidence of the information provided to decide whether the scenes shown are simultaneous, he argues; one simply uses the "brute fact" that they are all being shown simultaneously. Analogously

> We exploit [a] fact about the whole system [of self-conscious perceptual mechanisms, that they] are all wired together as the systems of a single animal. There is no need to use the specific contents of the information one has to ensure that it is a single thing in question.

He concludes that one knows that all the first-person knowledge one has concerns a single object.[66]

Campbell here trades on an equivocation concerning exploitation of knowledge and information. At first, he writes, our perceptual systems "[take] it that information obtained through different ... systems relates to the same time ... because [they] all produced their various pieces of information at the same time", then later it is "we [who] exploit [the] fact ... that the various perceptual systems are all wired together as the systems of a single animal". But the first consideration cannot have a

---

[65] O'Brien also poses the question in her (1995b), pp. 238–9: "How does a subject know that there is one object to which it is dispositionally related, to which it is identical?" Her slight concession to the possibility that the question may be self-defeating is to say that the capacity for self-identification requires an unexplained capacity for self-reference; see Chapter 1.

[66] Campbell (1994), pp. 100–1.

bearing on how *I* know that all my self-knowledge concerns a single subject; only if "exploit" means "knowingly exploit" could the fact that all the systems are wired together result in such knowledge.

The alleged exploitation of the fact that the perceptual systems are wired together is not comparable with the assumption that what seems to be closed-circuit TV screens are not using recorded material – which they might well be. Thus Campbell equivocates between personal and impersonal concepts of information. A more fundamental problem for him is that neither the "specific contents of the information", nor the fact that the various systems are wired together, can determine that there is a single thing in question, because the very use of "I" presupposes reference to a single subject.

With this we conclude the discussion of the unity of consciousness. The next chapter returns to Wittgenstein's original statement of IEM and the question of self-reference and self-identification, and examines issues which required a full delineation of IEM before they could be explored.

6
Self-Identification and Self-Reference

We now return to the question of self-reference and self-identification, and examine issues which required a full delineation of IEM: the relation between IEM and guaranteed reference of "I"; the opposition between direct and indirect reference theories of "I"; and whether there is genuine self-identification in IEM cases. Proponents of direct reference, who hold that the self-reference rule fully captures the meaning of "I", hold that the apparent redundancy of identifying knowledge shows that the reference of "I" is unmediated by a sense. Fregean proponents of indirect reference, in contrast, deny that grasp of the self-reference rule is sufficient to characterise self-consciousness, and require in addition apprehension of a sense for "I", which they must reconcile with apparent absence of identifying knowledge. This chapter attempts to undermine the debate between direct and indirect reference by drawing on Wittgenstein's "no-reference" view of "I", and arguing that IEM-judgements do not involve genuine self-identification.

A key contention of this chapter is that to view oneself as a person among persons, the requirement on self-consciousness defended in Chapter 1, does not imply a substantive or identifying self-conception. That is, one's self-recognitional capacities do not require a self-conception, or an accurate conception of the thing to which one's uses of "I" refer. The result is a deflationary account that provides further reasons for rejecting Cassam's "problem of misconception" criticised in Chapter 5. Cassam argues that perverse self-conceptions expressed by the avowals "I am a steam-engine" or "I am disembodied" pose a problem for the conceptual version of materialism concerning self-consciousness – hence his experiential version involving the self-presentation thesis. Here I argue that misconception is a pseudo-problem. A *two-tier model of self-consciousness*, involving both self-discrimination and self-non-discrimination, is

defended, suggesting that IEM-exhibiting capacities do not involve genuine self-identification. A *deflationary account of self-location* undermines the Fregean model, and we can see that there is no genuine identification of the subject in IEM-judgements – the truth in Wittgenstein's suggestion that "I" does not refer.

6.1 Guaranteed reference and IEM

Guaranteed reference is exhibited by all comprehending uses of "I". IEM, in contrast, applies only to a range of first-person utterances. Thus it is *terms* that exhibit guaranteed reference, and *judgements* that exhibit IEM. "Guaranteed reference" can mean: (i) "I" is immune from reference-failure: in using "I", a subject cannot fail to self-refer; or (ii) the intended and actual referent cannot fail to coincide. Only the first of these is unique to "I". In contrast, the necessary coincidence of intended and actual referent is not limited to "I"; on any construal of the distinction between speaker's and semantic reference, "I" and proper names or referential uses of definite descriptions behave similarly. If I look at an old photograph and say "I was wearing a funny hat", and speaker's reference is to the person in the photograph, there is no guarantee that speaker's and semantic reference of "I was wearing a funny hat" will coincide. The same applies to "The PM is wearing a funny hat", where the person observed wearing a funny hat is not the PM. If in the first case speaker's reference is to myself, since in using "I", I intend to refer to myself, then in the second case, speaker's reference is to the PM, since in using "PM", I intend to refer to the PM.

In contrast, in sense (i) – "immunity from reference-failure" – the behaviour of "I" is unique. "That" refers to the object which the speaker intends to indicate, if there is one; for "I" – and "now", if it is a referring term – there is no question of "if there is one". I am guaranteed success in referring to myself by means of "I" even when I lack current or recollected bodily information. Since guaranteed reference says simply that the existence of the referent is guaranteed, it can be regarded as a metaphysical truism: if there is an utterance there must be an utterer.

Many writers have conflated IEM and guaranteed reference in this sense, however. Anscombe writes that

> if "I" is a "referring expression" at all, it has both kinds of guaranteed reference. The object an "I"-user means by it must exist ... nor can he take the wrong object to be the object he means by "I". (The bishop

may take the lady's knee for his, but could he take the lady herself to be himself?)[1]

This is a mistake. Certainly the object must exist, but the second kind of "guaranteed reference" which Anscombe mentions can only be IEM. In non-IEM cases the "I"-user may take someone else to be himself; it might turn out to be someone else, and not myself, who is wearing the funny hat in the old photograph. Perhaps Anscombe's confusion arises from the Wittgensteinian argument that "no identification implies no reference": that genuine identification is required for reference, and that since misidentification is senseless, there is no genuine identification in IEM cases. (We return to this issue at the end of the chapter.) In fact, IEM presupposes guaranteed reference in the sense of immunity to reference-failure. IEM is the exclusion of the possibility that *I* – reference guaranteed – could acquire knowledge of *someone else's* past or body. No judgement whose subject-term lacked guaranteed reference could exhibit IEM.

6.2 Limited defence of indirect reference as not associated with observational model

Indirect reference theorists claim that self-conscious self-reference requires a "self-concept" or identifying conception. As noted earlier, Frege regarded the conception as a private one, or at least as involving private self-presentation: I am presented to myself "in a particular and primitive way" in which I am presented to no one else.[2] Anscombe shares Frege's intuitions about sense though not privacy, however, for Wittgensteinian reasons she concludes that they cannot be satisfied, and so advocates a "no-reference" view.[3] Indirect reference accounts and Anscombe's no-reference view of "I" are opposite sides of the

[1] Anscombe (1981b), p. 30. Brinck also seems to conflate IEM and guaranteed reference when she writes that "'I', at least in uses that are IEM, seems to be protected against reference-failure" (Brinck 1997, p. 92). But "I" is protected against reference-failure in all uses.

[2] Frege (1967), pp. 25–6. However, May (2006) comments: "Frege is *not* recommending private senses [for 'I']; to the extent that [he allows] the possibility of private ways of self-presentation, their overlap with one's ideas of oneself would render them unsuitable as the senses of 'I'. What he *does* recommend is a sense ... that constrains, but does not present, a reference ... The sense that Frege recommends for 'I' may not be private, but it is nevertheless egocentric" (pp. 502–3). On these issues see also Rodel (2007), Chapter 1.

[3] Discussed at the end of the chapter.

same Fregean coin, while direct reference theories lie in a different, non-Fregean plane. However, it is also true that in denying that "I" has a sense, Anscombe and the direct reference theorist share a deflationary account of the knowledge involved in self-conscious self-reference.

This chapter attempts to undermine the dispute between direct reference and Fregean accounts, but without adopting Anscombe's "no-reference" approach. The indirect reference view should be defended from the criticism that it is associated with an implausible demonstrative or observational account of the reference of "I", however. Exploring this criticism brings out contrasts between self-reference and self-identification. As noted in Chapter 1, it is commonly held that Anscombe's key error is to assume that if "I" is a referring expression, then it must be assimilated to the perceptual demonstratives. That is, she fails to consider the possibility that "I" belongs to a class of token-reflexive referring terms including "here" and "now". Rightly finding assimilation with the demonstratives implausible, the diagnosis concludes, Anscombe denies that "I" refers. We questioned this diagnosis, arguing that Anscombe explicitly rejects, rather than simply fails to consider, a token-reflexive account. For a Fregean, the token-reflexive option is unacceptable since it flouts the requirement that reference involves a mediating conception.

Although Evans describes Anscombe's non-referential conclusion as "extraordinary", he shares her Fregean assumption that reference implies identifying knowledge – an assumption that he calls Russell's Principle.[4] Evans's relative neglect of the self-reference rule, especially in his discussion of the parallels between "I" and "here", has led to the implausible suggestion that he wishes to assimilate "I" to the perceptual demonstratives, with the types of distinctively self-conscious self-knowledge playing the role that perception of an object does in the case of demonstratives.[5] This interpretation of Evans is unlikely. The following passage, quoted from earlier, suggest a more balanced reading:

> the essence of self-consciousness is self-reference, that is to say, thinking, by a subject of judgments, about himself, and hence, necessarily, about a subject of judgments… But it certainly does not

[4] Evans (1982), p. 212n.

[5] Brinck (1997), p. 39; O'Brien (1995b) both suggest this interpretation of Evans. They attribute to him the view that a reference-fixing information-link helps to build up a conception of the referent in both first-person and demonstrative cases; reference requires both the information-link and the conception.

> follow … that he must think of himself exclusively as an author of judgments, or even as a possessor of a mental life … our self-conscious thoughts about ourselves also rest upon various ways we have of gaining knowledge of ourselves as physical things.[6]

This does not suggest a demonstrative model of the *reference* of "I". Evans's reservations concerning the self-reference rule are not that it contradicts a demonstrative model of self-reference – for which he shows no sympathy – but that exclusive concentration on it encourages an unacceptably Cartesian conception of the subject as essentially thinking thing.

Evans assumes materialism about self-consciousness when he claims that the subject fundamentally identifies itself as a physical object: "a fundamental identification of a person involves a consideration of him as occupying such-and-such a spatio-temporal location … [thus] to know what it is for '*o* = I' to be true, for arbitrary *o*, is to know what is involved in locating oneself in a spatio-temporal map of the world".[7] In showing how this identification might be achieved, Evans may subscribe to an observational or demonstrative model of self-identification, but not of self-reference. However, it is not clear that Evans subscribes even to an observational model of self-identification. Responding to Anscombe's Tank Argument, he defends the more cautious thesis that the self-conscious subject must at least be disposed to have sensory information about itself.[8] One can agree with Evans that the requirement of a physical self-conception is an antidote to Cartesianism, without holding that it implies an observational model of self-identification.

Campbell associates an indirect reference account with a demonstrative model of self-reference. Though conceding that the demonstrative model really concerns self-identification and not reference-fixing, he still discerns a conflict between token-reflexive and demonstrative models. Although the self-reference rule fixes the reference of "I", he writes, "there is persistent pressure to suppose … that [a perceptual demonstrative] model must come back into play when we consider the relation of the first person to self-consciousness". Thus, he believes, there is a tension between how the reference of the first person is fixed, and the

---

[6] Evans (1982), p. 213. O'Brien considers that Evans might simply have been outlining preconditions of first-person thought, but reiterates that he is certainly aiming to provide a sense for "I" (1995b, p. 241).

[7] Evans (1982), pp. 210–11.

[8] Evans (1982), p. 216.

bases on which first-person judgements are made, and the consequences drawn from them – between reference-fixing and conceptual role.[9] On this view, the demonstrative model of self-reference is simply an assertion of the primacy of the conceptual role of "I" over the determination of reference.

To ask whether conceptual role is prior to reference-fixing is, I would argue, just to ask whether the self-reference rule presupposes IEM – which Chapter 1 argued it does. In order to understand the generality expressed by "each one" – as in "'I' is the word each one uses…" – one must be able to distinguish between oneself and others in a way that implies IEM. But the asymmetry reference-fixing of "I", and justification of judgements involving "I", seems unproblematic, and is exhibited by other terms. The reference of the subject-term of "That chair is made of cane" is fixed by what the judger is attending to and what concepts they subsume it under, while the basis of the judgement lies in visual perception. But the background conditions of reference-fixing are inseparable from the conditions for the comprehending use of "I".[10] My conclusion is therefore that an indirect reference account need not involve an observational model, and that Evans, in espousing indirect reference, does not advocate one.

6.3 A deflationary account replaces direct and indirect reference

I now argue that both direct and indirect reference accounts are mistaken. The data on which proponents of direct reference rest their case are largely an appearance, but an indirect reference account does not follow. The alleged self-conception should be characterised negatively in terms of an absence of deviant belief; there is no "problem of misconception". Cassam's solution – the self-presentation thesis – is an incoherent solution to a pseudo-problem.

First, to assess the case against indirect reference. We have seen that the indirect reference account is not associated with an implausible demonstrative model of self-reference. However, it still faces the objection, from proponents of direct reference, that a self-conception is not required for successful reference involving "I". In cases of bizarre delusion, successful reference using "I" appears to survive the almost complete disappearance

[9] Campbell (1994) pp. 96, 109.

[10] In particular, if they attempt to identify themselves in the circumstances in which Baldy does, the subject's grasp of "I" is called into question (Chapter 2).

of any associated conception or sortal. The same is not true of proper names or demonstratives.[11] More positively, while one can be inducted into use of the term "Oedipus" by being given certain true descriptions of Oedipus, there is no comparable introductory identifying knowledge for "I". "I" also differs from proper names and definite descriptions with respect to misinformation or lack of information. No amount of misinformation can cause failure of self-reference; "I" has a special propensity to stick with its object of reference through variation in the information about itself which the subject possesses.

Cassam argues, against objections to indirect reference arising from these features, that "the fact that 'I' is governed by the token-reflexive rule does not show that self-consciousness does not in any sense depend on one's having an accurate conception of the thing to which one's uses of 'I' refer", namely a material entity.[12] As we saw in Chapter 5, he believes that this requirement gives rise to a problem of misconception. In order to show that no accurate conception is required, we must examine two striking illustrations appealed to by direct reference theorists. Anscombe's *Tank Argument* seems to show that, even in the absence of current or recollected bodily information, I am guaranteed success in referring to myself by means of "I". First, however, we will look at the *Argument from Delusion*, which, according to Campbell, shows that "I"'s token-reflexivity refutes the indirect reference account. He argues that deluded subjects who believe that they are made of glass, that they are a steam locomotive, and so on, still successfully self-refer, hence a correct self-conception is not required for successful self-conscious self-reference.[13] Cassam responds that it is the sanity of such subjects, and not the requirement of an accurate self-conception, that should be questioned. Presumably he means that the deluded utterances of an insane subject are not comprehending. This debate should be undercut. Even if the utterances of the deluded subject are comprehending, a clear attribution of belief to them may not be possible – that is, they may not really believe what they say. First I will show why Cassam's suggestion

---

[11] As Anscombe shows (1981a), p. 28.

[12] Cassam (1997b), p. 646.

[13] Campbell (1994), p. 220. As Cassam notes, it is not clear that Campbell really does believe that a self-conception is not required, given his recognition that there must be a richness in the conceptual role of "I"; a different account is offered in Campbell (2001a). Cervantes's short story "The Glass Graduate", which was published in 1613 and came to the attention of Descartes, concerns a young man with the delusion that he is made of glass (Cervantes 1972).

of loss of understanding is not tenable, before defending the claim of no clear attribution of belief.

Loss of understanding is implausible because it would have to be local and not general – a possibility that makes no sense. The Capgras patient, who says that their partner has been replaced by a double, has not lost a general understanding of the terms they use; in other contexts they grasp what "my wife" and "that woman" refer to, and what "replaced by a double" means. Similarly, a schizophrenic patient who says that the arrangements of tables in a restaurant shows that the end of the world is coming, has been using the words "the world" and "ends" correctly most of their life. In the latter case a loss of understanding is even less plausible since it would have to be intermittent, occurring only during florid phases of the illness. Many – perhaps by definition all – bizarre delusions seem to exemplify what Ryle termed a category-mistake. "My internal organs have been removed", "I am made of glass", and "I am a pumpkin" are examples; "My wife has been replaced by an imposter" or "I am working for President Bush", in contrast, are not. However, the person who regards the university as a mysterious entity separate from its component faculties, or the mind as a mysterious entity separate from the body – to take Ryle's own examples – is meant to have a pervasive misunderstanding of "university" or "mind".[14] On the loss of understanding thesis, in contrast, the psychotic subject's misunderstanding is purely local; they understand "that woman" in most contexts except those featuring in the delusion itself.

The idea of a local category-mistake, and – it would follow – the loss of understanding thesis itself, may result from an over-literal interpretation of Wittgenstein's remarks in *On Certainty* concerning G. E. Moore's attempt to refute the sceptic.[15] Moore based his refutation on apparently commonsensical claims such as "I know that I have a hand", but Wittgenstein undercuts the debate between "common sense" philosopher and sceptic: "'I don't know if this is a hand'. But do you know what the word 'hand' means?"[16] That is, he suggests, anyone, whether sceptic or proponent of common sense, who attempts to debate whether this is a hand, thereby puts into question their understanding of the word "hand". Note that this is a sceptical doubt concerning a framework

[14] Ryle (1949), pp. 17–25 and *passim*. There much to be said on nonsense and understanding; there is a discussion in Hamilton (2013).

[15] Eilan in her (2001) treats these remarks as supporting the loss of understanding thesis as defended in Campbell (2001a).

[16] Wittgenstein (1969), para. 306.

proposition: a category into which many bizarre psychotic delusions fall, for instance the delusion that the inside of my skull is empty. Moore assumes that in entertaining the sceptical doubt he is saying something intelligible; by his response, "But do you know what the word 'hand' means?", Wittgenstein questions this. However, Wittgenstein is not claiming that it really is likely that a reasonably sane person such as Moore, in the course of philosophical discussion, has suffered a local loss of understanding of the word "hand". "You have not given your words a clear meaning" does not imply "You do not know what the word 'hand' means in this context"; the most that one could say is that it is *as if* the philosopher does not understand what "hand" means. Hence, although a *temporary* loss of understanding may perhaps be possible, there is no such thing as a *local* loss; one cannot make a local category-mistake.[17]

Instead of proposing a local category-mistake, therefore, one should recognise that there can be no clear ascription of belief in the case of delusions such as "I am made of glass", "I am a steam-locomotive", "I can pass straight through that closed door", or "I can be in two places at once".[18] It is not possible seriously to interpret someone as having such beliefs.[19] Certainly the more bizarre the delusion, the less likely it is to count as a belief – or, more precisely, the less likely that there is a fact of the matter concerning whether S believes that p. Monothematic, behaviourally inert delusions should not be regarded as beliefs, while polythematic, behaviourally active delusions may possibly be so regarded. (In monothematic delusion, it is apparently only a single belief that is affected, and the delusion tends to be behaviourally inert; in polythematic delusion, patients tend to be delusional about anything that attracts their attention.) Two features support the claim that delusions may not be genuine beliefs. The first is *groundlessness* – the notable feature of psychotic delusion that the patients offer no grounds for their delusions, or that if they do, they seem not to take them seriously. More mundane delusions – those which have the same propositional content as ordinary empirical beliefs based on

[17] It is interesting, however, that Wittgenstein does not favour the Humean view that the sceptic does not really doubt that the best way to leave the room is by the door, in preference to claiming that they do not understand what they are saying. The issue is discussed in Hamilton (2013), Chapter 11.

[18] The alternative thesis of a deviant understanding is even less plausible than that of a loss of understanding, as I argue in Hamilton (2007b).

[19] This line of argument does not undermine the authority of avowals of belief defended in Hamilton (2000), since, according to the definition of authority offered there, when the subject's avowal is not defect-free, the belief in question may not be attributed.

grounds – are distinguished by being ungrounded. The psychotic patient seems convinced that they are the Virgin Mary or whoever, but their justifications have the flavour of confabulation. Psychiatrists comment that "They don't mean anything to the patient"; they do not seem to be operative as reasons.[20] The self-deceiver recognises that evidence is applicable, but has a curiously selective way of interpreting it; the psychotic patient seems not to recognise that evidence is applicable at all.

The second feature that supports no clear attribution of belief is *behavioural inertia*. Many delusions exhibit relative behavioural inertia, and are not fully acted on. As Bleuler commented, although many schizophrenic subjects have delusions that they are great leaders, "None of our generals has ever attempted to act in accordance with his imaginary rank and station".[21] Similarly, sufferers from Capgras's delusion, which is relatively monothematic, often fail to express curiosity or form beliefs about where their spouse has gone, whether they are alive or dead, and so on. The delusion is avowed, and anger or irritation may be expressed towards the double, but often no attempt is made to search for the genuine spouse, and the subject may even be actively friendly towards the alleged double. When, in some cases, the delusion is acted on, sometimes with tragic consequences, it is closer to a belief. But even here there are often inconsistencies in accompanying feelings, and a curiously circumscribed quality to the delusion itself.[22] Jaspers contrasts the "specific schizophrenic incorrigibility" with the normal dogmatism of fanatics or of manic-depressives. The schizophrenic's delusion is quite unshakeable, he explains, yet in contrast to dogmatics, their attitude to the delusion is "peculiarly inconsequent at times", and such delusions often do not lead to action.[23] It is definitive of schizophrenia that the delusions are often not accompanied by an emotional state appropriate to their content; a schizophrenic patient may report that others are trying to kill them, while apparently remaining completely indifferent to this prospect.

These features of delusion suggest that no clear attribution of belief can be made in such cases. The resulting analysis of delusions parallels the analysis of self-deception which I have offered elsewhere.[24] The latter denies that the self-deceiver believes both that p and that not-p, where p is for instance "My partner is being unfaithful"; rather, there is evidence

[20] Comment to the author by Anthony David.
[21] Bleuler (1950), p. 129.
[22] As psychiatrists often note Young (2000) notes, p. 53.
[23] Jaspers (1963), p. 4.
[24] Hamilton (2000).

that they believe that p, and evidence that they believe that not-p. In self-deception, the conflict of evidence concerning whether X believes that p or whether they believe that not-p is *basic*; one should not assume a further, possibly evidence-transcending fact of the matter concerning their belief. The case of monothematic psychotic delusion, I now argue, is analogous; there is an equally irreducible conflict of evidence. There is evidence that X believes that p and evidence that they do not believe that p, while in self-deception there is evidence that X believes that p and evidence that they believe that not-p. The evidence for the ascription of the belief is that the subject, apparently without intending to deceive, repeatedly asserts p; the evidence against ascription is that they act on the putative belief at best in rather circumscribed ways, and seem not to hold it on the basis of reasons.

The claim is not quite that in many cases delusions are not beliefs, rather, it is that there is no fact of the matter concerning whether S believes that p. This position is subtly different from that of a number of recent writers who have argued that delusions should not be regarded as beliefs.[25] It also offers a less radical alternative to Cassam's suggestion that the deluded subject has lost understanding of the terms that they are using. Thus my conclusion is not quite that the patient does not believe that their partner has been replaced by a double. Rather, it is that there is evidence that the subject believes that p, and evidence that they do not believe that p, and that the conflict between these items of evidence is irresolvable. In everyday cases it is possible that even if X believes something firmly without it being any kind of delusion, there could still be evidence against the fact that they believe it. But our assumption is that the question concerning what X believes can be settled, and that any undecidability reflects X's own indecision. In the case of many delusions, in contrast, the question cannot be resolved. "What do they believe, then?" is the wrong question. There is a proposition concerning which there is evidence that the subject believes it, and evidence that they do not, and that is the best that can be said: a puzzling phenomenon. If one asks "What is the problem with this person?", three kinds of answer may be given:

(1) He believes that his wife has been replaced by a double.
(2) He claims, without intending to deceive, that his wife has been replaced by a double, but he does not believe it because he does not

[25] For instance Currie (2000); Graham and Stephens in Chung et al. (eds) (2007).

act in ways consistent with the belief, does not base it on grounds, and so on.

(3) He claims, without intending to deceive, that his wife has been replaced by a double, but we cannot say that he believes it because he does not act in ways consistent with the belief, does not base it on grounds, and so on. There is no fact of the matter about whether he believes it or not.

I am proposing the third answer. Underlying this conclusion is what may be termed an "*evidence is the base-level" thesis*, namely, there are cases of undecidable conflict of evidence which indicate that there is no deeper fact of the matter. It may be felt that the basic level of description in such cases cannot be evidential, since the very use of the term "evidence" points to a more basic level: – that which the evidence is evidence for. Clearly this is normally so. But the puzzling problem of self-deception and psychotic delusion is that the two sets of criteria for the ascription of belief, involving verbal and non-verbal behaviour respectively, do not cohere. Hence the undecidability.

The preceding analysis shows that in the case of many apparently problematic delusions, no deviant self-conception cannot clearly be attributed, and so Cassam's "problem of misconception" does not arise; there is no misconception. However, one might ask whether it arises in non-psychotic cases, such as the Cartesian avowal that I am an essentially immaterial entity. It is a familiar post-Humean response to the Cartesian or sceptic that they do not act in ways consistent with what they profess, and so do not really believe it. For instance, it may be suggested that someone who really does believe that they are an immaterial entity might attempt to pass through walls, yet those professing this belief never exhibit such behaviour. Cassam is unimpressed with this response. He argues that in the non-psychotic case, what people say provides the best evidence for what they believe, and, moreover, that "it is not obvious what kind of non-linguistic behaviour is supposed to manifest the Cartesian's (alleged) [common sense] belief that the subject of her thoughts is corporeal, as distinct from the [avowed, philosopher's] belief that her incorporeal thinking self 'has' a body".[26] The first point to be made in reply to Cassam is – to reiterate the conclusion just defended – that when linguistic and non-linguistic behaviour fail to cohere, beliefs cannot clearly be attributed, so it is not a matter of

[26] Cassam (1997a), p. 132.

better or worse evidence. In these problematic cases there is indeed a conflicting non-linguistic manifestation and so the required coherence is lacking. It is true that if, instead of "I am essentially disembodied", the Cartesian avowal takes the subtler form "I am lodged in my body like a pilot in a vessel" – a form which Descartes himself famously rejected, as we saw – then more care is needed in framing the conflict with non-linguistic behaviour. But such non-linguistic behaviour could include the subject's failure to treat their body as a prosthetic device, in the manner of Ian Waterman.

Anscombe's Tank Argument seems to strengthen Cartesian doubt because the subject, amnesiac but sane, currently receives no information from their body, and so apparently has a reason for doubt. As we saw, however, the having of sensory experience *presupposes* that I have a body, and so cannot provide evidence for it; so a lack of sensory experience in the Tank cannot justify a doubt about whether I have a body. Epistemically the situation is no different from that of conventional Cartesian doubt. Nonetheless, it is difficult to see how the subject in the Tank could behave in ways that conflict with the doubt that they have a body, and so it seems that linguistic and non-linguistic behaviour are consistent. This is not so, however. The almost complete attrition of any sortal conception is only an appearance, as shown by the consideration presented in Chapter 1, that in grasping the self-reference rule, the subject must understand what "each one" – as specified in the rule "'I' is the word that each one uses…" – refers to. Thus even if the self-reference rule is regarded as the essence of "I", the understanding of "each one" at least involves viewing oneself as a member of a community of language-users or persons; a community of disembodied subjects seems imponderable. If this understanding of "each one" is lost in the Tank – as *ex hypothesi* the subject's Anscombian doubt suggests – then no "I"-thought could be attributed.

The preceding considerations undermine the direct reference standpoint, I will argue, without thereby establishing the Fregean one. As argued in Chapter 5, "I am embodied" is a problematic claim, for there is no intelligible alternative to conceiving or experiencing oneself as embodied; the Strawsonian requirement says that I regard myself as a person among persons, rather than as "embodied". "I am embodied" or "I have a body" are much more problematic propositions than "I am a human being" or "I am a person", though these are not unproblematic, since it might also be argued that there is no alternative to being a person among persons. A Strawsonian sortal conception is not properly a "self-concept", which would imply the contemporary Fregean structure

of writers such as Evans and McDowell, who add an individual concept or de re sense to the basic self-reference rule.[27] This stable self-concept is meant to address theoretical problems such as providing the cognitive significance of "I", displaying the self-knowledge a reasoning agent must have, and integrating conceptual and non-conceptual content; proponents maintain that it underlies IEM, and gives rise to locating beliefs necessary for agency.[28] Rejecting this complex self-conception, I now defend the deflationary proposal that there can be no genuine ascription of beliefs to the subject which is inconsistent with the belief that they are a person among persons.

6.4 Deflationary account developed: discrimination and non-discrimination requirements

A deflationary account undermines the expansive Fregean "self-concept" that includes capacities to locate and discriminate oneself. The anodyne truth in the Fregean claim is that I know, in an everyday sense, which object I am: knowledge about my past, my spatial location, and my relations to others. Evans regards this as ordinary descriptive knowledge, and maintains that it presupposes a fundamental, non-descriptive self-location and self-identification. As we saw, that involves considering the person "as occupying ... a spatio-temporal location ... to know what it is for '*o* = I' to be true, for arbitrary *o*, is to know what is involved in locating oneself in a spatio-temporal map of the world". Evans believes that such identification is essential to self-conscious thought, and that it depends on the various IEM-exhibiting capacities of self-location, orientation, and proprioception, as well as perception.[29] For Evans, the claim of fundamental identification flows from Russell's Principle: that one cannot think about anything, including oneself, without being able to discriminate it from the rest of the world:

> a subject's self-conscious thought about himself must be informed (or must at least be liable to be informed) by information which the

[27] Evans (1982), chapters 1, 6, 7; MacDowell (1984; 1986). Brinck has developed this account, arguing that the individual concept has two components: a de re sense with close links to the context of utterance, and a fundamental self-concept stable through different contexts, consisting in an intuition of oneself as a subject of thought and experience (Brinck 1997, pp. 109–21).

[28] These are the roles claimed by Brinck (1997), p. 140.

[29] Evans (1982), pp. 210–11. The claim of fundamental self-identification is also defended by O'Brien (1995b).

> subject may gain of himself in each of a range of ways of gaining [such] knowledge ... and at the same time the subject must *know which* object it is of which he thus has, or is capable of having, knowledge.

Cassam stipulates a similar condition, derived from what he calls the *discrimination requirement* that the ascription of different states to an identical subject depends on some means of identifying the subject of such ascriptions as one object among others.[30]

I wish to deflate these claims without endorsing Nagel's radical suggestion that the subject does not really know what it is for such identity-propositions as "I am AH" to be true, that there is an unbridgeable gulf between subjective and objective. A similarly radical position is adopted by Anscombe, whose no-reference view of "I" means that she also denies that "I am AH" is an identity proposition.[31] The Evans-Cassam position requires qualification, not rejection, in light of the Wittgensteinian thought that in a central range of cases – those exhibiting IEM – it makes no sense for the subject to identify or discriminate which person they are. Hence I propose to supplement Cassam's requirement with the *non-discrimination requirement*: I cannot think about anything without there being something, myself, which in central cases I do not need to discriminate.

On the most radical interpretation, arising from a no-reference view of "I" and the metaphysics of the disappearing self, the non-discrimination requirement *excludes* the possibility of self-discrimination. A more moderate compatabilist position holds that the discrimination requirement is satisfied in non-IEM self-ascriptions, while the non-discrimination requirement is satisfied in IEM-ascriptions. The result is a *two-tier model of self-consciousness* that qualifies Evans's claim that self-identification depends on IEM-exhibiting capacities, since these are just the ones which do not – at least directly – involve self-identification.

The discrimination requirement follows from the standard assumption that "I" refers to an object, and that objects must be discriminable. My

---

[30] Evans (1982), p. 212; Cassam (1997a), p. 121. Cassam refers to Strawson's statement, in his (1966) p. 102, of the "discrimination requirement".

[31] Nagel (1970) and (1986), *passim*; Anscombe (1975). In fact, central uses of "I am AH" include giving one's name – meaning "I am called AH" or "My name is AH" – and in these uses it is not an identity proposition. Evans regards Anscombe and Nagel as advocating an "Idealist" conception of the self; Anscombe's positive account, however, which says that "I"-thoughts are made true by the behaviour of this living human being, of which I have a special, unmediated knowledge, means she may not really be an "Idealist" at all.

concern is not with the non-discrimination requirement as such, but with showing that IEM-judgements involve no self-discrimination. However, the non-discrimination requirement is plausible because identification cannot be indefinite, on pain of an infinite regress. If, alternatively, one holds that identification is like belief in involving a judgement – it is active, not passive – one could argue that Evans's position threatens an explosion of acts of self-identification that is as unacceptable as an explosion of beliefs.[32] A deeper reason why I do not identify myself in central cases may be that I am the locus of my own actions, thoughts, and experiences, including my acts of self-identification.

Discrimination and non-discrimination requirements are not essentially in tension. That I do not need to discriminate something, and indeed must be able to proceed without in fact discriminating it, does not show that I cannot discriminate it – nor that my being able to discriminate it is not necessary. "Necessarily sometimes I do not do X" is compatible with "Necessarily always I am able to do X". However, according to the position defended here, the non-discrimination requirement is satisfied by self-ascriptions where the idea of self-discrimination makes no sense. On this interpretation, non-discrimination is incompatible with the claim that one can discriminate oneself in other ways. So although the two requirements as such are consistent, they may conflict in how they are satisfied. Satisfying the non-discrimination requirement implies a weaker interpretation of self-discrimination than Evans specifies.

However, the two requirements are most convincingly reconciled by holding that while IEM-judgements do not involve an actual discrimination, they presuppose a general capacity of self-discrimination. It seems plausible to say that in order to make IEM-judgements one must also be able to make non-IEM-judgements, and vice versa. The denial that IEM-judgements involve genuine self-identifications rests on Wittgensteinian arguments concerning epistemic guarantees. In non-IEM cases, there is the possibility of ignorance and error; I may believe that someone is F without realising that it is myself, or I may believe that I am F when in fact it is someone else. Thus I may identify, mis-identify, or fail to identify myself as the boy at the end of the back row in the school photograph; as the person who caused the trail of sugar through the supermarket; as the killer of Laius, which in Oedipus' case involved identifying as Laius the man he had killed. None of these self-identifications is IEM. Since identifications in this category assume that the subject is

[32] This issue is pursued in section 4 below. The explosion of beliefs was criticised in Hamilton (2000).

a person, they cannot themselves be the basis for such a fundamental identification – nor do they confirm or provide evidence for it.

What underlies everyday identifications, such as identifying my youthful self in the school photograph, is an anodyne truth in the Fregean claim. In the photograph, I look for a spotty youth with long hair and glasses. Underlying the identification is my knowledge of my own appearance, and, more fundamentally, my knowledge that I have a past, that I looked and still look like a human being – though these are not really "knowledge", but rather part of a *world-picture* in the sense suggested in Wittgenstein's *On Certainty*. It is the individual's world-picture, if anything, that underlies everyday self-identification – a notion distinctively different, in form if not content, from G. E. Moore's notion of a common-sense view of the world, the original inspiration of Wittgenstein's thinking. Characteristically, it is not an explicit object of the individual's propositional thinking, but rather, constitutive of their bodily life.[33] To postulate a fundamental identification, in contrast, is to attempt to respond to a pseudo-problem.

6.5 Deflationary account concluded: self-location

A deflationary analysis of the self-concept must also address self-location, in particular undermining the assumption that self-location involves fundamental identification. Distinguishing the varieties of self-location shows why Evans's claim that a subject locates itself fundamentally in a spatio-temporal world map involves a misleading metaphor. His contrast between ordinary descriptive self-identification and fundamental self-identification glosses over a range of self-locating capacities, including what I will term *habitual*, *indexically descriptive*, and *mapping self-location*, and *egocentric* and *allocentric* frames of reference. None of these is fundamental, absolute, or objective in Evans's sense.

Evans wrongly privileges *mapping self-location*, which is neither primitive nor fundamental, and does not involve an absolute, as opposed to allocentric, self-location. Many societies do not have maps, and as Jerry Brotton writes, "…for thousands of years… [people] did not think of [maps] as being in a category separate from the writing of formal documents, painting, drawing or inscribing diagrams on a range of different media from rock to paper".[34] Indeed, there is a problem even in

[33] Wittgenstein's notion is developed in Hamilton (forthcoming 2013).
[34] Brotton (2012), p. 5.

defining what maps are. Harley and Woodward suggest "graphic representations that facilitate a spatial understanding of things, concepts, conditions, processes, or events in the human world".[35] We can say that aerial photos are not maps, though they could be just as helpful, and in the same kind of way, in navigating. An aerial sketch might look even more like a map, without being one – because unlike a map it does not depend on conventions, for instance that # is a power station. One does not need to know what a power station looks like to recognise one on the map, as one does in order to recognise one in a photo or sketch. But there is at least a vague boundary between a map and diagram; consider a more schematic version of the already schematic Frank Beck London Underground map. Furthermore, as Brotton also argues, "…the mapmaker's physical observation of the earth…is never made from a neutral cultural standpoint…the map is always a creative interpretation of the space it claims to represent".[36] He has in mind, for instance, different methods of global projection, which traditionally overemphasised the northern rather than the southern hemisphere.

Thus mapping is one of the less primitive kinds of self-location. Nor is it fundamental. I locate myself when through habitual knowledge I find my way to a destination such as home; when I can direct someone indexically; or when I know where I am on a map. The last of these abilities should not be privileged. If I ask passers-by in Rome how to get to the Coliseum, one person may identify the monument, and our present location, on a street map; another may give directions – "Straight on, turn right, second left…"; a third cannot give verbal instructions but walks with me to the monument. The second person's guidance involves indexical description, and is action-guiding for a subject who cannot read maps or has no street plan. In this and the third case the subject knows where they are without knowing their location on a "map of the world". "They know how to get there" means "If you ask them they can tell you", or "They can find it without any help".

Indeed, one can know where one is on a map, without being able to find one's way about in the territory covered by it: "knowing where I am" in the practical sense. "Knowing where I am on a map" may describe (a) pointing to the correct place on the map and saying, "I am here", or (b) saying, correctly, "I am at Little Puddlecombe". These are different capacities, and neither of them implies the practical ability. Thus "I know where I am (on the map), but have no idea how to get to

[35] Harley and Woodward (eds) (1987), p. xvi.
[36] Brotton (2012), pp. 6, 14.

anywhere from here" is not an absurd statement. One cannot assume that someone who knows where they are on a map will know how to use it.

These examples bring out the obvious fact – maybe not so obvious that it does not need stating – that knowing where I am means knowing where other things are too, in the practical sense of being able to find my way to them. I know where I am in Rome, for instance, if I can get to the Coliseum and other places from my present location. "I know where I am, but I've no idea how to get to anywhere from here", except where it means "I know where I am on the map" *is* an absurd statement. Indeed, habitual self-location often amounts simply to destination-location, because unlike mapping self-location, it is non-inferential. A good striker in football is one who seems to know instinctively where the goal is, even when spun round or knocked over and therefore apparently disoriented. One could say that they are always able to orient themselves; their surroundings are always "familiar". Using a map, in contrast, involves self-location as well as destination-location. Mapping self-location is a more intellectual capacity; it is genuine self-location and not merely homing. But it is not thereby more fundamental; rather, it presupposes the other kinds of self- or destination-location.

Evans espouses a *skill theory* of egocentric space, which says that a creature grasps its content in possessing and using a range of sensorimotor skills (though its claims are too simple to amount to a "theory").[37] It holds that knowing where I am is essentially a matter of finding my surroundings familiar: knowing how to get to places in the vicinity, in the indexical sense. This is in tension with Evans' assumption that mapping self-location is fundamental, which reflects the contemporary consensus arising from Tolman's work on "cognitive maps". These are defined as mental representations analogous to topographic maps, which – it is said – enable a creature to navigate its environment, "computing" the shortest distance between two points.[38] This popular invocation of inner representations is both intellectualist and pseudo-explanatory. Indexical and habitual self-location do not involve an implicit mapping. Knowing where the Coliseum is – or where the fridge is – is a practical ability, which someone possesses if they can find it.

Cognitive or mental maps in Tolman's sense were a breakthrough in social or perceptual geography, but they do not explain in the way that conventional maps do. They may be explanatory, but not as usually

[37] The "theory" is discussed further by Grush (2000), p. 64.
[38] Tolman (1948).

assumed, which regards them as internal maps that enable the subject to locate themselves and navigate to places. The subject consults no such map, any more than a pigeon consults a compass. But the geographical observer can construct cognitive maps to represent, and predict, the subject's navigational behaviour. For instance, a Londoner's "mental map" of Scotland – in conjunction with their use, if any, of real maps – will explain their route-planning.

Mapping abilities are therefore rather sophisticated, and rely on habitual and indexical self-locating capacities. Maps are just one way of enabling people to get to places, and many societies do not understand them, just as some, it is said, do not understand pictures; "evidence for the level of map consciousness in early societies ... is virtually nonexistent".[39] J. J. Gibson's ambulatory account of perception argues that cognitive maps as internal entities are unnecessary, given the abundance of perceptual information concerning the temporarily hidden.[40] My argument is that even if an explanation of self-location were required, cognitive maps could not supply it, since mapping self-location, implicit or not, assumes more fundamental kinds of self-location.

The claim of fundamental identification is further undermined by another important distinction. Mapping self-location is *allocentric* – literally "centred on another" – while indexical and habitual self-location is *egocentric*. The egocentricity of self-location is paralleled by that of spatial perception, which is articulated by means of indexical terms relating objects to the perceiver – "over there", "up here", "to the right".[41] An egocentric frame is held to be defined by axes left, right, up, down, front, behind, with an origin identified as "here". Indexical self-location involves such descriptions as "the statue is at the first turning on the right", "the painting is over there", and "Louise is just here". The subject is the frame of reference – the origin of the space within which other things are located – to which these indexical terms refer. Their body and behaviour define axial asymmetries – up or down, left or right, ahead or behind. In contrast to egocentric self-location, however, in *allocentric or*

[39] Harley and Woodward (eds) (1987), p. 508.

[40] Gibson holds that psychologists committed to the retinal image and snapshot theory of vision have failed to distinguish such perceptual information from the genuinely imaginary. Reed (1988), pp. 300–3, discusses Gibson's largely unpublished views on this topic.

[41] The term "egocentric" is used by Evans and Campbell. For instance, in "Molyneux's Question" in his (1985), pp. 387–8, Evans discusses the relation of egocentric spatial perception to conscious experience.

"*other-centred*" *self-location* the frame is provided by some object other than the subject.

Some varieties of egocentric space may have a non-ego reference point; one might say that the statue is to the left of Jones, while still using an egocentric frame. But where the object has no natural orientation or asymmetries, one should regard it as the origin of an allocentric space centred on, and with axes provided by, the object itself.[42] Virtual points of view are also allocentric: an example would be a diagram of four objects in a room – myself perhaps one of them – as seen from a point ten metres above. Dead-reckoning systems which enable a creature to locate itself by keeping track of how fast it has been moving, in what direction, and for how long, are allocentric; an animal may use the position of the sun, plus its knowledge of the time of day, as a compass, and plot its route home. However, as Campbell points out, the animal has to know the direction home not just in terms of the external "compass"; it also has to know which way to point itself to travel in that direction. He concludes, plausibly, that egocentric axes are those immediately used in the direction of action, and that grasp of egocentric frames of reference must be taken as primitive: "[It cannot be defined] in terms of the generic notion of an object-centred frame of reference plus the notion of a body-centred frame".[43]

Campbell contrasts "egocentric" and "absolute" frames of reference and conceptions of space, but the term "absolute" is misleading. Against Campbell, and Evans also, I maintain that egocentric and allocentric viewpoints exhaust the possibilities – there is no such thing as an absolute or objective viewpoint. As various writers have argued, the tendency to think of maps as involving such a viewpoint is mistaken.[44] Maps are just another kind of allocentric representation of space, constructed relative to human interests, in order to prioritise desired objectives or conspicuous features – as the abstraction from parallax illustrates.[45] It is grasp of egocentric space, not the mapping viewpoint, which is absolute, fundamental, and primitive.

---

[42] As Grush (2000) comments.

[43] Campbell (1993), p. 75; Campbell (1994), p. 12. In this, I believe, he echoes claims made more obscurely by Merleau-Ponty in his (1962), Part 1, Chapter 3.

[44] For instance Grush (2000).

[45] Grush comments: "Most maps are somewhat stylized versions of such a representation, abstracting away from particularities that create parallax, for instance. But they still have an orientational structure, left-right and up-down axes. They are identical to the 'object centered' reference frames, except that there happens to be no actual object or person that is constituting the point of view" (Grush 2000, p. 80).

This is shown both by the fact that egocentric axes are those used immediately in directing action, but also in that grasp of egocentric space arises together with the ability to locate one's own body-parts.[46] I know where my fingers are if I can move them; in a different sense, I know where they are if I know that they are in the trifle (allocentric space). Ian Waterman's limbs are not paralysed, but move wantonly and uncontrollably. He does not know immediately where his arm is, not because he does not know that it is resting on the cocktail cabinet – a normal subject in a darkened room might not know this – but because he cannot move or control it. Thus the ability to locate one's limbs and other body-parts *consists* essentially in the ability to act, and does not *explain* it, as proponents of a Fregean self-concept maintain. Appeal to inner body-maps is no more explanatory than appeal to internal maps of one's environment. The mapping of efferent nerve endings onto brain structures reflects topological features of the body, but does not constitute a mental map. To produce a robot one might need to implant such a map, but it does not follow that self-conscious subjects actually possess one.

It could be argued that the two-tier model of self-consciousness is not really a model of self-consciousness because, while higher-level mapping ability is distinctive of self-conscious creatures, habitual self-location is a capacity shared with non-self-conscious creatures. Certainly, indexically descriptive and mapping self-location are restricted to self-conscious creatures, since they involve linguistic and other high-level capacities; but one should be cautious in attributing self-locating abilities to non-self-conscious creatures. Capacities towards the homing end of the spectrum cannot be described as genuinely self-locating. *Homing ability*, manifested by pigeons, and by the territorial behaviour of cats, is the most basic and primitive. It is illustrated also by Herdwick sheep in the English Lake District, who are "hefted" to their own fell and do not stray from it; culling during the foot- and mouth-crisis of the late 1990s might have meant losing this acquired knowledge for generations. Even if some self-locating abilities are shared with non-self-conscious creatures, habitual and indexical self-location remain more fundamental for self-conscious creatures than mapping self-location. It is exhibited by them in a much more sophisticated, varied, and refined sense, over a larger physical area.

---

[46] This is Campbell's view. He claims that grasp of egocentric spatial axes must be taken as primitive, though he interprets the ability to locate one's body-parts in terms of the concept of a body-image that was criticised in Chapter 4 (Campbell 1993, pp. 72–4).

6.6 Two senses of "identification" and the no-reference view of "I"

This chapter concludes by returning to the origins of IEM in Wittgenstein's *Blue Book*. Earlier commentators have focused on Wittgenstein's inference from IEM to a "no-reference" view of "I", rather than on how IEM is a hallmark of self-conscious ways of knowing. But although it is not my main concern, some discussion of the former issue is required. Many writers would share Evans's reasons – flowing from his general account of reference and identification – for denying that IEM-judgements involve genuine identification. Although, as discussed earlier in this chapter, he does not fully endorse an observational model of self-identification, his mistaken assumption that IEM has a uniform explanation across the demonstratives and indexicals does lend it support.[47] IEM, he writes, "is a straightforward consequence of demonstrative identification ... [which] is, precisely, a way in which a thought can concern (be about) an object", and so Wittgenstein's move from the senselessness of misidentification to the denial of identification was "just a mistake":

> In one sense, anyone who thinks about an object identifies that object (in thought): this is the sense involved in the use I have just made of the phrase "demonstrative identification". It is quite another matter ... for the thought to involve an identification component – for the thought to be identification-dependent. There is a danger of moving from the fact that there is no identification in the latter sense (that no criteria of recognition are brought to bear, and so forth) to the conclusion that there is no identification in the former sense.[48]

I believe that Wittgenstein's move is not "just a mistake". He fully intended to deny that "identification in thought" means anything distinct from identification-dependence in the sense of applying criteria – and indeed to deny any sense of "refers" distinct from identification-dependence. Moreover, Evans's claim that IEM is uniformly explained as "straightforwardly a consequence of demonstrative identification" is undermined by his observation that demonstrative IEM is more limited than that of "I".

For Wittgenstein, identification is essentially an active matter, and, without a logical possibility of failure, one cannot speak of successful action.

---

[47] As Brinck points out in her (1997), p. 40.

[48] Evans (1982), p. 218. Wittgenstein's move is found in the well-known passage in the *Blue Book* discussed in Chapter 2.

On his view, since one could never *mis*identify the subject of first-person IEM-judgements, one cannot be said to *identify* them either. His argument has the form "If there is no logical possibility of being wrong, then there is nothing that counts as being right either".[49] The implicit requirement is that for a subject to be credited with some achievement, in this case a cognitive one, it must be possible for them to fail in it. Thus it would be absurd to suggest a race with only one "competitor"; they could not be said to "have won" it.[50] The requirement, succinctly, says "No wrongs do not make a right". Wittgenstein, of course, does not call it a requirement, or even state it explicitly, but it is a key idea in his later thinking, giving rise for instance to his discussion of whether I can know that I am in pain. It also underlies a central tenet of *On Certainty*, that claims to knowledge make sense only where doubt is possible – that "knowledge implies the logical possibility of doubt". The tenet first appears in *Philosophical Investigations*, Part II, p. 221: "'I know…' may mean 'I do not doubt…' but does not mean that the words 'I doubt…' are *senseless*, that doubt is logically excluded". Wittgenstein's implication is quite the reverse: "I know that p" requires that "I doubt whether p" has sense.[51]

In his *Blue Book* remarks on "I"-as-subject, Wittgenstein seems to argue the further step that since there is no genuine identification in IEM cases – that it makes no sense to talk of "recognising a person" – "I" cannot be said to refer in such cases. That is, he suggests a *no-reference account of "I"*, according to which it is not the function of the first-person pronoun to refer:

> "But surely the word 'I' in the mouth of a man refers to the man who says it; it points to himself; and very often a man who says it actually points to himself with his finger". But it was quite superfluous to point to himself. He might just as well only have raised his hand…[52]

---

[49] Shoemaker also endorses the idea that "identification necessarily goes together with the possibility of misidentification" (1968, p. 562).

[50] The slogan underlies the discussion of sensation "S" in the course of the Private Language Argument (PI 258–70), while a closely related principle is implied in the later discussion of intention: "But in the sense in which I cannot fail to will, I cannot try to will either" (PI 618). An analogous argument proceeds from the criterionless and incorrigible self-ascription of avowable states to the non-assertoric thesis of avowals. These issues are discussed in Hamilton (forthcoming 2013), Chapter 10.

[51] These issues are discussed in Hamilton (forthcoming 2013).

[52] Wittgenstein (1969), pp. 67. For some reason, he italicises the "I" in the "use as subject" examples.

Wittgenstein's endorsement of a no-reference view of "I" is not unequivocal, however, especially in his later period. In *Philosophical Investigations* he comments that, in using "I", I do not "name a person", while in a manuscript he writes: "It is correct, although paradoxical, to say: 'I' does not refer to a person". His fire, however, is directed more at the treatment of "I" as an object, "the self".[53] Wittgenstein's remarks were developed by Anscombe into an explicit no-reference account, though, as noted earlier, she mistakenly conflates guaranteed reference and IEM.

The no-reference account should not be equated with Hume's position, that "I" refers to nothing just in the way that "the present king of France" refers to nothing – that "I" is in the denoting business, but a failure at it.[54] For proponents of a no-reference account, "I" is not in the denoting business at all, and refers to nothing, just as "it" in "It is raining" refers to nothing. Here, "it" is a grammatical place filler, not a conventional way of indicating where it is raining or that it is raining here. No proponent of the "no-reference" view regards *all* uses of "I" as non-referring; many argue that only IEM uses are non-referring. "I'm going", when an expression of a desire to go, is an IEM use; in unusual circumstances, such as when I have consulted a list of those who are coming and misread someone else's name as my own, the utterance might not be IEM. Proponents of no-reference would have to say that the latter cases are not those to which one would appeal in explaining the use of "I" – but the asymmetry between uses creates further problems concerning truth-value links.

One version of the no-reference view treats "I" as a signalling device, which shows but does not state who is speaking.[55] On this version, the self-reference principle might be amended to: "'I' is the word each one uses to signal the speaker". "I" is assimilated to literal signalling behaviour, as when a teacher asks "Who's going on the trip?", and instead of saying "I am!", the children put up their hands. An alternative strategy

---

[53] Wittgenstein (1958), para. 404; MS 116, 215.

[54] As Van Inwagen neatly puts it, in his (2001). The "no-reference" view follows from Cassam's Exclusion thesis – that the thinking, experiencing self is not an object among others in the world – while the Humean view corresponds to his elusiveness thesis (Cassam 1997a, *passim*). There is a critical discussion of the "no-reference" view in O'Brien (2007), Chapter 2.

[55] "I am here" is redundant as an utterance except as a means of signalling to someone that I am in their vicinity; when pointing to my location on a map; or as a reminder to others of my presence, equivalent to "You're forgetting about me".

might be to separate *reference* from *denotation*: a term denotes, while a person refers using a term. Thus "Brandon" on a road sign denotes the village, while reference proper is an intentional action and therefore subject to the Wittgensteinian "no cognitive achievement" condition discussed above. For various reasons, however – the most pressing, again, concerns the existence of truth-value links – this is not a convincing strategy.[56]

It therefore remains preferable to separate reference as a passive grammatical notion from identification, rather than to distinguish reference and denotation. The result would be a distinction between semantic and speaker's reference, and rejection of semantic identification. Thus while reference might be said to involve no immediate cognitive achievement – although it presupposes a general linguistic competence that is partly cognitive – identification always involves a particular cognitive achievement. "I" always refers, but some uses (non-IEM) are identificational while others (IEM) are not. It follows that there are cases of reference – those involving "I" at least – where identifying knowledge is not deployed, or not directly deployed.

This is the conclusion that I wish to draw.[57] To vindicate it, one would need to show that thinking about an object is not something that just happens to the subject, and that it makes sense to intend to think of an object and yet fail to do so. Wittgenstein's view that reference essentially involves active identification seems implausible. If one construes object-direction in thought as the correlate of reference, then Evans is right to maintain that reference just is identification, understood in a passive sense. But Wittgenstein's view that there is no genuine identification in IEM-judgements is defensible, I believe. In order to discredit his claim, one must give an account of an "active" identification that can never fail, and such attempts are unconvincing, I believe. Thus I would argue that the truth in the no-reference view is the claim that there is no genuine identification in first-person IEM-judgements.[58]

This volume began with an outline of Wittgenstein's view of the importance of IEM. The no-reference view of "I" is too radical a conclusion to draw from the phenomenon, I have argued. The analysis of conceptual

---

[56] It is discussed in Hamilton (1987).

[57] The no-reference view of "I" was considered at some length in Hamilton (1991). For an ingenious recent defence of the view, see Wiseman (2010).

[58] Hacker (2001) also seems to conflate identification and reference when he writes that we fail to see that "the use of 'I' standardly involves no identification at all and at best only a degenerate form of reference" (p. 35).

holisms involving memory and proprioception proves a more fruitful development of it. Whatever view one takes on this question, however, philosophical work on self-consciousness must surely place IEM at the centre of its concerns.

The final chapter turns to the standpoint of *philosophical humanism*, which regards self-consciousness as essentially a human phenomenon, extended only in a primitive sense to certain higher primates such as chimps. It suggests further directions of enquiry, both into the nature of philosophical humanism, and into its implications for philosophical treatment of animal intelligence.

7
Humanism and Animal Self-Consciousness

The previous chapter, in developing the concept of self-identification, considered some issues of animal self-consciousness in terms of self-location. This chapter considers animal self-consciousness directly. It returns to the Analytic Principle discussed in Chapter 1 – that self-consciousness is a phenomenon that must be expressed by use of a self-referring device with the properties of the first person – and asks whether it is consistent with primitive self-consciousness in animals. If chimps and other non-language-users exhibit self-consciousness, then the latter capacity may appear detachable from "I"-use – hence the challenge to the Analytic Principle, which rests on the assumption that self-consciousness necessarily involves linguistic expression.

It will become clear in the course of this chapter that the implications of animal self-consciousness for the Analytic Principle are not at all straightforward, however. This is because the nature of animal thought is itself a deeply intractable philosophical problem. Since the body of empirical work on animals and self-consciousness is huge as well as interesting, it is not possible in this volume to engage with it at appropriate length, and this final chapter is an open-ended prelude to future work on the topic. It has the modest aim of suggesting some lines of enquiry, and exploring possible responses – focusing on the *mirror test* which seems to illustrate self-consciousness in chimps. However, I believe that we can conclude that when primitive self-consciousness is understood as non-basic – as derivative from full self-consciousness – the Analytic Principle is consistent with its attribution.

This chapter begins by outlining a position called *philosophical humanism*, which excludes the possibility of animal self-consciousness in anything other than a derivative or fragmentary sense. Philosophical humanism affirms the importance of humane understanding against

both scientism and – a less common view nowadays – supernaturalism. This chapter shows various ways in which the humanist position may be developed. Although this volume's central arguments stand independently of the case for humanism, its treatment of self-consciousness and personal identity may be helpfully elucidated in terms of the humanism debate.

To reiterate, the Analytic Principle assumed by this monograph rejects a fundamental level of self-conscious or "self-specifying" thought not involving use of the first person. The present chapter develops this claim, arguing that since higher primates do not attain the level of first-person thought, they exhibit self-consciousness only in a primitive or fragmentary sense. The result is a humanistic standpoint that contrasts with recent work of a naturalistic or scientistic hue, such as Bermúdez'.[1] As we saw in Chapter 1, he claims both that the human case is paradigm, and that there is a fundamental phenomenon of self-consciousness – exhibited by non-language-using creatures – that does not involve use of the first-person, and which is explained in evolutionary terms.[2] The resulting position is, I believe, incompatible with philosophical humanism. Rather than regarding primitive self-consciousness as underlying all others, humanism treats the richer human self-consciousness as basic, with other kinds derivative of the human paradigm. On this view, only rudimentary self-consciousness should be imputed to chimps.

7.1 Philosophical humanism in the philosophy of mind

Before turning to animal self-consciousness, the key features of humanism debate will be addressed, in order to explain my use of the term "humanism". Since its everyday usage is somewhat loose – in contrast to "scientism" – my aim in this first section is a critical reconstruction and sharper definition of the concept. I am concerned both to affirm my allegiance to an already extant position, and to outline more sharply that position's commitments. This chapter therefore involves a creative, and not a merely scholarly, exercise.[3]

In its original meaning, "humanism" referred to the learned culture underpinning the European Renaissance. Scholars or *umanisti* revived

[1] Bermúdez (1998); other examples are Carruthers (2000); Chalmers (1996); Dretske (1995); Rowlands (2007).

[2] Bermúdez (1998).

[3] To take a somewhat lofty parallel, this is what Mill does in *On Liberty* with liberalism, though he does not yet use the term.

the study of classical philosophy, literature, history, and law, and – according to the traditional grand narrative – the resulting impact of Ancient Greek and Roman culture and civilisation transformed the medieval consciousness. For Cicero and other ancient authors, *studia humanitatis* was a liberal education centred on authoritative Greek and Latin texts, comprising grammar, rhetoric, and poetics. Under the same title, that relatively unified group of disciplines were studied in Italian universities from the fifteenth century, partially displacing Aristotelian scholasticism. For Copenhaver and Schmitt, "Humanism was not a field of learning in its own right but a method, a style, and a curriculum that various disciplines found useful. There were medical, legal and mathematical humanists as well as philosophical humanists".[4] These scholars revived the learning of Greek and Latin antiquity; early humanists such as Petrarch regarded the medieval era, as we now call it, as barbarous and uncultivated.[5]

As is often the case with cultural, religious, and historical concepts, *humanism* is a much later historian's term for this learned culture, first appearing as the German *Humanismus* in the later eighteenth or early nineteenth century.[6] The French *humanisme*, or "love of humanity", appeared contemporaneously, and eventually became the central meaning – a secular world view which rejects supernaturalist or religious standpoints, and affirms the centrality of the human realm. Copenhaver and Schmitt comment that humanism in some contexts "connotes an aggressive anthropocentric secularism quite foreign to the Christian world of early modern Europe".[7] Philosophical humanism, as I am characterising it, is not aggressively secular, but stands between two positions. It is defined primarily in opposition to *scientism*: the view that the physical or natural sciences constitute the paradigm of human knowledge, on which other disciplines must model themselves. However, it also rejects the (normally religious) view that I term *exceptionalism,* which holds that "human animal" is a contradiction in terms,

---

[4] Copenhaver and Schmitt (1992), p. 36.

[5] The "grand narrative" expressed in this paragraph has been challenged by recent historians, notably Wickham (2010).

[6] A useful account of these developments is found in Dear (2001), Chapter 2. Baldwin (2007) discusses the origins of the term "humanism". His primary concern is the debate between Heidegger and Sartre concerning humanism as a metaphysical commitment to truth as representation, underlying which is the humanist "privileging" of the human subject, targeted by post-Structuralist and Marxist critics as well as Heidegger.

[7] Copenhaver and Schmitt (1992), p. 24.

and that human beings are the only biological entity that cannot be grouped with others on any level. The opposed poles that humanism rejects, then, are scientism and exceptionalism.

The physical or natural sciences principally comprise physics, chemistry, and biology, sciences that deal with the objects, phenomena and laws of nature and the physical world. The term "physical sciences" emphasise that scientism, in most of its manifestations, is committed to the foundational status of physics. Contemporary philosophy of science and philosophy of mind debate how to define physicalism. To refer to current physics is too restrictive, but "future physics", it is argued, might include psychological states, on the assumption that a neurophysiological reduction of them is possible. "Natural sciences" is an alternative term, under which psychology might also be included; psychology and the social sciences are not, for scientism, appropriately foundational. Mathematics is another discipline whose scientific status is unclear. But even if its formulation is problematic, scientism as a loosely defined tendency clearly exists.

Philosophical humanism opposes scientism through both philosophical and methodological claims:

(1) The explanation of human behaviour is irreducibly personal, that is, it essentially involves what is often termed *the intentional stance*, resting on common-sense psychology and the attribution of beliefs, desires, intentions, and similar attitudes. Explanation of human behaviour, on this view, is not exclusively personal; subpersonal and neural explanation has a place, but not, as in scientism, the ultimate one. The fact that it accords subpersonal and neural explanation a place means that humanism does not amount to exceptionalism as defined above, whose implications will shortly be pursued.
(2) The humanities comprise a relatively distinct group of disciplines concerned with the human rather than physical world, whose forms of explanation are not reducible to those of the natural sciences.

It may also embrace a third claim about the discipline of philosophy itself, though as we will see this claim is not essential to humanism:

(3) Philosophy is assimilated with the humanities, and is not a proto- or meta-science, though it overlaps with the sciences, and with religion.

Scientism denies all three of these claims. (There could perhaps be a version of scientism that did not regard philosophy as a proto- or

meta-science, but as a unique discipline not conforming to a scientific paradigm. Humanists who deny (3), in contrast, assume a three-way classification of humanities, sciences, and philosophy.)

Bernard Williams and Peter Hacker advocate humanism in the tripartite sense just outlined, and both stress methodological claims (2) and (3).[8] Hacker argues that the humanities require concepts not deployed by natural sciences. To describe and explain animal behaviour, he argues, zoology requires only an attenuated range of the psychological concepts employed by the humanities, a range restricted to what is expressible in the animal's limited behavioural repertoire.[9] Bernard Williams, treating philosophy as part of a "wider humanistic enterprise of making sense of ourselves and our activities", rightly takes history as the paradigm humanistic discipline, and his views will be considered further below.[10] Hacker makes the essential point that one can acknowledge the value of science without being committed to scientism: "Forms of rational understanding and explanation are diverse and logically heterogeneous. Science and humanism were ... allied in [combating] unreason, moral and political dogma, and ... religion".[11] Humanists must acknowledge the value of science; not to do so would be absurd. Their opposition is to a philosophical misreading of science.

Claim (1) requires considerable elucidation. Rejecting scientistic philosophies of mind, it says that whole-person ascription involving the intentional stance is the fundamental level of explanation of human behaviour. Humanism does not regard subpersonal or brain processes as irrelevant to the attribution of psychological states and explanation of human behaviour. Consideration of the subpersonal level is often required; the humanist claim is that human behaviour cannot be completely, or best, explained in terms of it. An instance would be the treatment, and explanation, of depression; the effectiveness of drug therapy clearly involves subpersonal explanation, while that of psychotherapy may not. The problem, in comparing the two lines of treatment, is how to connect these two levels of explanation. But what humanism insists is that explanation of human behaviour essentially involves whole-person ascription.

---

[8] Williams (2008), Hacker (2001). Another recent critic of scientism is Charles Taylor.

[9] Hacker (2001), p. 69.

[10] Williams (2008), p. 197.

[11] Hacker (2001), pp. 71–2. By "humanism" he means "the humanities".

Self-consciousness, on the humanist view, is paradigmatically human, and attributed to non-humans insofar as they resemble the human case. This humanist position underlies the Analytic Principle, with its insistence on language-use as essential to self-consciousness, illustrated in particular by the defence of a personal concept of information in Chapter 2, and by the present chapter's discussion of proto-self-consciousness. Humanism also rejects the uncritical use of science-fictional thought-experiments, such as Parfit's reductionist attempt to show that physical animal and human person are not coextensive. Such thought-experiments require careful handling, and their results are often indeterminate, as we saw in connection with q-memory in Chapter 2. Using them without regard to the human paradigm, as Parfit does in his notion of person-fission, constitutes ungrounded speculation that runs counter to humane understanding.

What is the "human paradigm", exactly? As we will now see, humanism denies that "human" is essentially a biological concept, scientistically construed, and holds instead that it refers essentially to a rational, cultural, language-using creature. It therefore rejects *animalism*, the currently popular deflation of the personal identity debate, which argues that we are animals of a certain human kind – essentially, human organisms. This view treats "human" as a biological notion, a position rejected by humanism.[12] "Human", in the humanist's book, does not simply have the biological meaning of "pertaining to homo sapiens" or "the genus homo". It also implies personhood, where a person is a rational, cultural, language-using creature; animalists, on this view, are wrong to doubt that humans are essentially persons. For instance, when we say "Kipper is such a clever dog, he is almost human", we do not mean that biologically, despite appearances, he is close to Homo sapiens.

A defensible philosophical humanism, therefore, privileges not the biological species "Homo sapiens" – as those who convict it of "species-ism" allege – but human beings as sentient, rational, language-using, and culture-producing creatures. This is the version of humanism that I am advocating, whose commitments I am concerned to interrogate. It underlies Bernard Williams's humanistic, anti-biologistic response to evolution: "it is not ... human cultural practices that are explained by natural selection, but rather the universal human characteristics of having cultural practices, and human beings' capacity to do so. It is precisely the fact that variations and developments in

[12] On animalism, see for instance Olson (1997); Snowdon (1998); and Noonan (2010).

cultural practices are *not* determined at an evolutionary level that makes the human characteristic of living under culture such an extraordinary evolutionary success".[13] To say that "we are essentially human animals", biologically construed, is to say that the human characteristic of living under culture is inessential to what "we" are. This, to the humanist, is not credible. Ontology must, whatever its proponents say, be conditioned by an implicit direction of interest. The humanist argues simply that the term "human", as it appears in "human animal", implies other directions of interest and dimensions of meaning in addition to the biological ones acknowledged by the animalist.

Having contrasted humanism with the scientistic standpoint of biologism, it is essential not to assimilate it with *exceptionalism*. I develop this term from Dupré's reference to "traditional exceptionalist accounts of our species", accounts that reject scientistic standpoints such as evolutionary psychology, and which advocate a kind of "species-ism".[14] The term "exceptionalism" is normally used to capture myths of national uniqueness; American, Israeli, or English exceptionalism are largely illusions, or at least inflated pretensions. Humanism is not exceptionalist, because it is not based on a myth of uniqueness; there is no myth, nor any claim of uniqueness. Exceptionalism is the position that – relying on one common use of "human", which contrasts it with "animal" – regards "human animal" as an oxymoron. It regards human beings as too special to be grouped with any other part of the animal creation. When fundamentalist Christians reject Darwinism on the grounds that they find it repellent that we might be descended from monkeys, they are advocating an emotive form of exceptionalism. For humanists, in contrast, the development of Darwinian theory is a great humanist success.

"Human" is standardly used in biology to refer to the genus Homo and not just to the species Homo sapiens. Recent theories suggest that the Neanderthals, who have had a bad press, were culturally as well as biologically "human": they buried their dead, and made weapons, which are cultural activities.[15] If Neanderthals had continued to evolve in tandem with Homo sapiens, and lived alongside them, we ("Homo sapiens") would presumably have called them persons. It is true, however, that we have no experience of the coexistence of intelligent species, and,

---

[13] Williams (2008), p. 188.
[14] Dupré (2004), p. 53.
[15] Perhaps the biological sense of "human" is a genus term, the genus "Homo", but my arguments against animalism remain.

indeed, evidence only for the reverse; dominant species see off other species, as Homo sapiens probably killed off the Neanderthals.

It may be objected, against the humanistic claim that language-users are the paradigm of self-consciousness, and that an animal's behaviour must be understood by assessing how far its cognitive powers approximate to human ones, that "language-user" – or "rational" or "cultural being" – and "human" are not coextensive. There are severely cognitively impaired humans who cannot use language, and perhaps there are non-human language-users. I will first address the category of *cognitively-impaired humans*. Proponents of this objection, I believe, are wrong to deny that language-use – or cultural production or rationality – is a universal human ability, in the sense of being a part of normal human development. Either the human individual is developing towards it, or for some pathological reason, has lost or never had it.

These Aristotelian or teleological considerations, which form the beginnings of a response to the objection, are often surprisingly neglected – for instance by Colin McGinn, when he considers the claim that "Marginal humans [the very young, senile or brain-damaged] differ morally from animals … because they are members of a species whose *typical* members are full persons". This position, he writes, "locates the basis of a being's moral rights not in [its] intrinsic nature … but in its relation to other beings … if the human species manages so to pollute the earth that in time a typical member is no longer a full person, almost everyone having been chemically brain-damaged, then no human will enjoy moral significance".[16] His use of the term "damaged" shows that McGinn retains a conception of the healthy, undamaged individual; this is the criterion, however, not for statistical typicality, as his account implies, but for normality. A normal human is not like a typical car with automatic drive (which uses more fuel, for instance) or a typical Durham student (who comes from a private school).[17] These merely statistical facts are unconnected with purpose, function or nature.

To turn now to the objection arising from *non-human language-users*. To reiterate, I am concerned to reject two opposed poles: scientism and exceptionalism. Humanism stands between these positions. My conclusion is that "human animal" is not a strictly biological notion. Animalists gain support for their position by implicitly trading on an equivocation between two senses of "human", with biological and non-biological

[16] McGinn (1996), p. 40.

[17] This Aristotelian line of thinking is found in such writers as Lear (1988), Chapter 5, and Megone (1998).

implications. My opposed account does not deny that we are animals, and is therefore not exceptionalist. It allows that culture is not autonomous, and interacts with biology.[18] Although one should be sceptical about the claims made for primates as language-users, philosophical humanism allows that Koko the gorilla, if she counts as a language-user, could exhibit self-consciousness in something like a full as opposed to primitive sense. This fact shows that the humanism advocated here is neither dogmatically exceptionalist, nor biologistic; it allows that although the human case is paradigm, it is not unique. Insofar as other, non-human "species" of person became more commonly known to us, that paradigm status would decline.[19]

Koko, and the Neanderthals, are, crucially, animals. There is something about our notion of person that is grounded in being animal and therefore in part, but not exclusively, investigated by biology (however defined). Animalism's error is to claim exclusivity for biology as the science of the personal. "Human" is only in part a biological concept. A fact of deep contingency – that almost all the persons with which we are acquainted, are Homo sapiens – means it is not just a biological concept. Humanism does not, therefore, simply dismiss the possibility of alien societies at least as sophisticated as our own. It says, in Wittgensteinian vein, that if alien societies were discovered, our concept of personhood would change. The humanist claim that self-consciousness involves a rational, cultural, language-using standpoint is an elucidation, though not a proof, of the Analytic Principle.

7.2 Philosophical humanism and the humanities

To return to methodological claim (2). The humanities – to continue the story of Renaissance humanism – are a group of disciplines that evolved from *studia humanitatis*. These include history, theology, classics, law, and the theoretical study of the arts, including literature, drama, music,

---

[18] Diet, for instance, is mixture of cultural and biological factors. The Japanese raw-fish diet is facilitated by bacteria in their gut that enables digestion; explanation of the prevalence of diabetes among indigenous peoples in Australia is both cultural and biological.

[19] Humanism therefore differs in emphasis from Baker (2000), which makes no reference to that paradigm status: "To be a person – whether God, an angel, a human person, or a Martian person – one must have the capacity for a first-person perspective. Person is a genus, of which there may be several species: human, divine, bionic, Martian, and so on. It is in virtue of the capacity to have a first-person perspective that a person is a person" (p. 92).

and visual art and architecture. In the nineteenth century, a debate arose from the growing sense of division between arts and sciences, concerning what, if any, distinctive methods they have. Philosophical humanists have often attributed to the humanities a method of understanding, or *Verstehen*, as opposed to causal explanation.[20]

There is a qualified truth in this view, but the picture is more fragmented and pluralistic. One factor is the highly contested place of the social sciences in the opposition between philosophical humanism and scientism. There is a debate, for instance – nowadays totally one-sided – between those who regard psychology as an experimental science, and those who believe that it should be a humanistic "science of the soul" as advocated by William James.[21] Proponents of scientism, while they perhaps include psychology and the social sciences within the realm of the sciences, would not regard them as paradigms of knowledge like the "hard" or physical sciences.

In contrast to the apparently modest support enjoyed by humanism, scientism is assumed by many contemporary philosophers, often unreflectively – a list of adherents would be very extensive. The opposition between philosophical humanism and scientism developed only during the course of the nineteenth century, when sciences and humanities began to be regarded as distinct. Scientism became fully formed in the later nineteenth and twentieth century positivist doctrine of the unity of science, and more recently in Quine's conception of philosophy as continuous, in method and subject matter, with science. The unity of science today is still defined in terms of both method and content, but rests in particular on the doctrine of the completeness of physics. As well as Quine, proponents include Goodman, and – in one philosophical incarnation – Hilary Putnam, who writes that "the philosopher has to…be a good 'futurist' – [to] anticipate for us how science will solve our philosophical problems".[22]

It could be argued that the science-humanities contrast is itself artificial, a relic of the very scientism I am concerned to reject.[23] Dupré, for instance, argues that the sciences are unified neither by method – no account of which has improved on Popper's unsatisfactory treatment – nor by content, since the doctrine of the completeness of physics lacks

---

[20] See for instance Von Wright (1971).

[21] The question is discussed by Reed (1997), who advocates a Jamesian humanistic psychology.

[22] Putnam (1989), p. 107.

[23] This might be another way of putting the argument of note 5.

empirical support. For Dupré, there is no such thing as scientific method, or the method of the humanities, nor are there two grand cultures of arts and sciences – but rather, many small and overlapping subcultures:

> If science is no more than an overlapping collection of practices for investigating the world, with diverse assumptions and methods… then it is unclear why anything much hangs on the claim to scientificity….

The resulting pluralism has the virtue that "complex phenomena are far more likely to be understood if a variety of distinct but complementary approaches are brought to bear on them…".[24] However, even if one cannot talk of two grand cultures, pluralists should concede that there are two broad tendencies, consistent with a heterogeneity of explanation and understanding.

The sociological or cultural divide between arts and sciences persists, to the detriment of both sides, and rests on an arts–science continuum of practices. Modern experimental psychology, for instance, clearly aspires to the scientific end of this continuum. And even if the arts and sciences form a more heterogeneous collection than scientism allows, the latter's erroneous conception of science has been widely influential in philosophy.

I conclude this brief outline of philosophical humanism by considering claim (3): that philosophy is one of the humanities. It is advocated by some Anglo-American philosophers, notably Bernard Williams, who takes history as the paradigm humanistic discipline, and holds that philosophical understanding of our ideas and motivations requires historical understanding. Conceptual analysis, he argues, is not self-sufficient, and the project of deriving concepts a priori and ahistorically from universal conditions of human life leaves many philosophical issues unexplained. Williams recognises that while epistemic and moral values such as accuracy and sincerity are universal, their forms vary historically and culturally: "If one is to understand our own view of such things [philosophically] one must try to understand why they take certain forms here rather than others, and one can only do that with the help of history".[25]

---

[24] Dupré (2004), pp. 53–7. Snow (1993) is the *locus classicus* for discussion of "Two Cultures", arguing that a common culture has been replaced by two distinct cultures, represented by scientists and "literary intellectuals" respectively. A recent discussion is found in Kagan (2009).

[25] Williams (2008), p. 192.

A humanistic standpoint affirms a close relation between philosophy and its history – and between philosophy and history as such. A powerful humanistic case can be mounted that while the history of science is history, not science, the history of philosophy is itself part of philosophy.[26] The other humanities – theology, law, politics, and the arts – align themselves here with philosophy. Concern with its own history has been an overt feature of philosophy since the Renaissance, and was implicit in philosophical practice as far back as Plato and Aristotle, who "tried to overcome their predecessors by describing and then rebutting them; in this sense, philosophy had historical motivations from the first…" However, medieval philosophers lacked historical curiosity, taking their understanding of the historical development of Greek philosophy uncritically from Aristotle. The first books titled "history of philosophy" appeared in the seventeenth century, by which time thinkers had developed the distinction between ancient and medieval periods of Western philosophy.[27]

Reading the thinkers of the past gives us philosophical insight. But the value of history of philosophy is more than this. Philosophy's concern with its own history helps to make the alien ideas of the past more familiar, allowing us to grasp the hidden world view of our own age – those assumptions too close to us easily to be visible – thus defamiliarising it, and allowing us to regard our own ideas as products of a moment in history. Thus philosophy's engagement with history goes far beyond concern with its own history.

The contrast between history and the sciences can be expressed by means of the concept of *vindicatory explanation*, through which a later

[26] This is not entirely correct; Ernst Mach believed that misunderstanding of science's history encouraged scientific error, and knowledge of the history of science may be valuable for scientists. The sciences are part of culture and therefore in a sense humanities (see for instance Knight (2009), Introduction). Perhaps mathematics has more of a humanistic affinity to its history than do the natural sciences.

[27] Copenhaver and Schmitt (1992), pp. 332, 331, 337. They comment that "The Renaissance, so seldom remembered by modern philosophers…first gave their discipline a continuous and purposeful sense of history…until [then] the surviving materials [of ancient philosophy] lay scattered and unexamined among the larger body of texts…of classical literature. Not just in philosophy but in all other respects, modern historical thought began in the Renaissance" (p. 332). Their description of Russell's famous history of Western philosophy is marvellously accurate: "a belletristic book written in a mood of affable aristocratic omniscience, an engaging amateur's history of the type that used to earn an author praise as a 'Renaissance man'" (p. 339).

theory explains the one that it supplants, allowing proponents of the older theory to come to recognise this process as an advance. As Bernard Williams comments, vindicatory explanation is commonplace in science, but not, despite the advocacy of Hegel and Marx, in philosophy.[28] If it were, the history of philosophy might properly be a side issue for those concerned with the pursuit of philosophical truth. But it is not. Philosophical standpoints are abandoned generally because of changing intellectual fashion, tastes, or commitments, rather than through refutation. With science, in contrast, refutation is more common; there is no philosophical equivalent of the refutation of phlogiston theory in the late eighteenth and early nineteenth century, and its replacement by an understanding of the process of oxidation. Debate between Platonists and Aristotelians is perennial; that between phlogistonians and oxygenists, or even Newtonians and Einsteinians, is not. Newton's theories are included in Einstein's. There are no alternative scientific traditions – Boyle-Lavoisier versus Davy-Faraday, for instance – and the enterprise is much more collaborative.

Williams possibly overstates the case, however. Philosophy is both an abstract space of theories, and a human activity – as indeed is science. Kant's work is commonly and rightly held to constitute philosophical progress, synthesising the truth in both empiricism and rationalism – for instance in replacing the theory of ideas with a theory of judgement – just as Einstein includes the truth in Newton. In showing why earlier thinkers went wrong, his philosophy, therefore, offers vindicatory explanation. There are later but essentially pre-Kantian philosophers who still made important contributions – notably Mill and later empiricists. But the reappearance of the theory of ideas in Russell's work – his mature philosophy has been described as Hume plus modern logic – was a regression.

Since philosophers are mostly not historians – and historians are mostly not philosophers – the historical process by which philosophical ideas develop has been neglected. Their causes are recognised only in the most general terms:[29] for instance in how from the seventeenth century onwards, philosophers have been impressed with, and responded to, the development of natural science; how, with the weakening hold of religion in Western societies, philosophical concern with religious ethics also declined; and how, relatedly, the post-Enlightenment growth of

[28] Williams (2008). Marxist historicism blurred the distinction between philosophy and science, therefore, even if Engels created "scientific materialism".

[29] The work of Jonathan Israel, notably his (2008), is a notable exception.

reason and decline of traditional authority led philosophers to develop accounts of moral and aesthetic autonomy.

Because philosophical explanation is generally non-vindicatory, philosophy reserves a special place for classic texts, even if they lack the status of their artistic equivalents. P. N. Furbank rejects the

> prejudice that eighteenth-century readers were naive and that we know better now… profound and utterly original literary masterpieces of the period, such as Diderot's *Le Neveu de Rameau*, Voltaire's *Candide*, [and] Rousseau's *Confessions*, [are] works which mankind will never grow out of.[30]

The same may be said of Hume's *Treatise* and Kant's *Critiques*, which do not simply contribute to our understanding of their own times, but which yield an enduring understanding of ourselves and our intellectual, imaginative, and moral capacities.

However, one can be a philosophical humanist as defined above while denying that philosophy is one of the humanities, endorsing (1) and (2), but rejecting (3). This version of humanism denies that philosophy is continuous with the sciences, but also denies it any special affinities with history and the humanities. This is the Analytic standpoint, and Wittgenstein's throughout his career:

> Philosophers constantly see the method of science before their eyes, and are irresistibly tempted to ask and to answer questions in the way science does. This tendency is the real source of metaphysics and leads philosophers into complete darkness.[31]

On his view, philosophy is a completely autonomous discipline, not to be assimilated to the sciences, social sciences, or humanities. Wittgenstein's rejection of claim (3) goes with his apparent lack of interest in the history of philosophy, and with his rhetoric about having founded a new subject, heir to what used to be called philosophy.[32]

Wittgenstein's conception of philosophy was therapeutic, but most of those who espouse his combination of views (1), (2), and (3) advocate the anti-historicist standpoint that I term *Analytic purism*. An anti-historicist standpoint is least persuasive in areas of philosophy

---

[30] Furbank (2003), p. 68.
[31] Wittgenstein (1969), p. 18.
[32] Ibid., p. 28.

whose key concepts have undergone great historical change, such as ethics, aesthetics, and political philosophy; in analysing the concept of art, for instance, analytic aesthetics is reluctant fully to acknowledge the implications of its radical historical developments. Even in philosophy of mind, historical considerations are significant. The Greeks lacked both a concept of art in the modern, post-eighteenth century Western sense, and a post-Cartesian concept of mind, while the personal identity debate did not take shape until the post-Cartesian era.[33] Such historical developments constitute important material for assessing contemporary common-sense psychology.

However, this volume's primary concern has not been with the historical development of the concept of self-consciousness.[34] The Greeks had minds, however they conceptualised the mind, and exhibited self-consciousness, however they conceived of it. In contrast, lack of a system of the arts had a profound impact on the presence of art in Greek society, even though we now regard some of their products as artworks. It makes no sense to speculate when self-consciousness appeared historically, except to say that it appeared with language, and with human society. Adorno and Horkheimer's discussion of the historical emergence of subjectivity, as the subject attempted to protect itself from a threatening natural environment through language, is intriguing but unpersuasive.[35] The main target of a philosophical humanist account of self-consciousness, therefore, is scientism that accords the subpersonal a central role over the personal, rather than scientism that neglects the historical conditioning of concepts.

Philosophical humanism requires a monograph in itself, and the preceding is only a sketch of its commitments. The rest of this chapter restricts itself to the question of non-human exhibition of self-consciousness. The interpretation of the mirror test for chimps offered here does not rest on humanism, but it is an illustration of it.

---

[33] The philosophical implications of the Greek concept of art and craft are discussed in Hamilton (2007), Chapter 1.

[34] On this more historical question, see for instance Martin (2004).

[35] For Adorno and Horkheimer in *Dialectic of Enlightenment*, Odysseus' encounter with Polyphemus illustrates the subject–object divide that expresses estrangement from the natural world (Adorno and Horkheimer 1987). When asked by Polyphemus for his name, Odysseus uses its double meaning to make the giant believe that he is no one, ironically affirming his identity as a subject by denying it.

7.3 The mirror test as illustration of proto-self-consciousness in non-language-users

We now continue with that question concerning self-consciousness on which humanism has a close bearing: the exhibition of self-consciousness by intelligent non-human, non-language-users. To reiterate, many hold that chimps and other higher mammals exhibit self-consciousness in at least a primitive sense; some will conclude that self-consciousness is detachable from "I"-use, and thus that the Analytic Principle should be rejected. This line of argument is misguided, I believe. The self-consciousness ascribed to chimps is a minimal or primitive *proto-self-consciousness* that depends on the richer, human, linguistic paradigm. To reiterate, proto-self-consciousness does *not* mean "basic and fundamental self-consciousness", as postulated for instance by Bermúdez.[36] The Analytic Principle is therefore consistent with the presence of proto-self-consciousness, in this sense, in creatures with at best rudimentary linguistic abilities.

Case studies of proto-self-consciousness focus, understandably, on self-recognition. Though difficult to test for in the absence of language, recognition of one's appearance does seem to be conspicuously exhibited by chimpanzees. The *mirror test*, which has inspired much psychological research, seems strikingly to illustrate this primitively self-conscious behaviour. Gallup showed that once chimps have become familiar with mirrors, they recognise their own mirror – image, as do dolphins and elephants as subsequent research has shown.[37] Chimps begin to use mirrors to look at previously unseen parts of their body such as the inside of the mouth; most interestingly, when a spot of paint has been placed on their forehead, they notice this in the mirror and rub it. Once the chimp has become familiar with mirrors, reaction to the spot is spontaneous and untrained. (I say "has become familiar", but reacting in this way to the spot may itself be part of the process of familiarisation.) The chimps were not rewarded for touching the spot; those without prior exposure to mirrors did not touch it. Monkeys, in contrast, treat their mirror images as they do other monkeys, usually by threatening them. Human children

[36] Bermúdez (1998).

[37] This research, conducted in 1970, is discussed in Hauser (2001), Chapter 5, pp. 122ff.; Cheney and Seyfarth (1990), especially p. 242; and Byrne (1995), pp. 111–17. Dolphins are discussed by Derbyshire (2001), and elephants by Highfield (accessed 2006). Typically, of those chimps tested, fewer than half pass the test.

between one and two years begin to pass the mirror test, reacting with surprise and delight and putting their hands to their foreheads.[38]

A famous recipient of the mirror test was Koko the gorilla, who was brought up much like a human, and taught American sign language:

> the researchers [wiped] Koko's brow with a cloth. She got used to this gentle and unthreatening manoeuvre ... [When] a harmless water-based paint was hidden on the cloth, and a large coloured mark thus placed above her eyebrows ... she caught sight of herself in a mirror – and jumped in comical "double take", proceeding to investigate the mark while carefully looking at her reflection ... she had already discovered a black pigment spot in her mouth with the aid of a mirror, and would readily reply "Me, Koko" in sign language if asked what her reflection was.

These reports make it appear that Koko was an "I"-user.[39] But my principal concern is not with the controversial attribution of language-use to intelligent animals, but with the rather less sophisticated responses of ordinary chimps.

Analogous tests for the sense of hearing, such as a chimp's response to a tape recording of its own calls, are possible. Tests on day-old human infants showed that they cried significantly less when played a recording of their own crying than one of the crying of other infants of the same age. For Bermúdez, these data show that the infants discriminate their own crying from that of others of the same age, and are capable of self-perception even at one day old.[40] This is too much to claim, I believe. The infant's response should be assimilated with being soothed by its mother's recorded voice or heartbeat: merely a preference for the familiar,

---

[38] See for instance Hauser (2001), p. 118. Dogs and cats exhibit a third kind of reaction, complete indifference: "Domestic kittens and puppies first threaten the 'animal' in the mirror, then try and investigate by pawing behind the glass, and gradually come to avoid or ignore mirrors altogether" (Byrne 1995, p. 112).

[39] Byrne (1995), Chapter 8. Compare Rogers and Kaplan, who say of a talking bonobo that his "ability to understand requests improves when they are made in syntactically complete sentences, compared to truncated pidgin English". See Dupré (2002), Chapter 11. An animal with such a linguistic grasp would be a non-human person, but provided such cases were comparatively unusual, they would be consistent with humanism.

[40] Bermúdez, describing the research, regards this behaviour as "a clear case of self-perception" (1998), p. 125.

An elephant approaches the jumbo-sized mirror

not involving the judgement "That is how I sound" or "That is how my mother sounds". The baby's crying reaction is much less sophisticated than the chimp's in the mirror test, so there is less inclination even to interpret it as having "I"-thoughts.

What is the significance of the mirror test, and why is it assumed to illustrate primitive self-consciousness? On the humanist view proposed here, self-consciousness is not a unitary phenomenon across species, and only a rudimentary form should be imputed to chimps who pass the test. Furthermore, the test is not the only evidence for proto-self-consciousness in higher primates. Chimps live in quite complex social groups whose hierarchies are based on sex and strength; their relationships are usually of a dominant/submissive nature. They carefully observe each other's eye movements, which gives rise to another example of behaviour expressive of proto-self-consciousness: female chimps position themselves behind a rock or other obstruction so that the dominant male cannot see that a younger male is grooming them. Proto-self-consciousness is also shown in strategies of deception; a member of a troop of chimps, feeding in a group, notices some bananas, looks round to check that the others are oblivious, then returns later to gorge itself.[41] Note that although it is able to deceive, a chimp may not

[41] Primatologist Jane Goodall observed a chimpanzee in the wild, that she had named Figan. On seeing a banana in a tree above a more dominant chimp, Goliath, he moved away and sat down facing away from the banana, to avoid

be able to think of its fellows as able to deceive in their turn; for this it would have to possess the *concept* of deception.[42]

The chimp therefore exhibits proto-self-consciousness through a range of intelligent behaviour. These capacities – recognising itself in a mirror, deception, tracking the gaze of others, planning strategies of becoming dominant, and of gaining help in these strategies – are primitive self-consciousness in some of its forms and manifestations, which characteristically appear in a set and co-vary. That is, creatures tend to have all or none. There is, I believe, no deeper fact of the matter concerning the presence of self-consciousness, than the presence of something like this range of capacities; self-consciousness is not something underlying these capacities. The chimp's mirror-behaviour goes with other proto-self-conscious behaviour; conversely, the monkeys' failure at mirror self-recognition is one of various behaviours indicating a less complex intelligence not involving proto-self-consciousness.

These points apply a fortiori to the capacities of human beings. People who lived before mirrors appeared – though glass began to be used in the Middle Ages, use of shiny metals is very ancient – were self-conscious even if they could not initially have recognised themselves in a mirror. (Reflections in water have presumably been available to people and other creatures throughout the ages, however.) The New Guinea tribe whose Stone Age culture did not include familiarity with pictures or photographs reacted with fear and incomprehension to the "people in the glass" until mirrors were explained to them, yet they were undoubtedly self-conscious.[43] An individual has to belong to a society familiar with such artefacts before one can say that failure of mirror or photographic self-recognition implies either lack of self-consciousness or cognitive impairment.

---

alerting the other chimp, Goodall argues. When Goliath left shortly after, Figan climbed the tree and ate the banana. The example's implications are discussed in Singer (1993), p. 116.

[42] Lowe (2000), p. 181, and Wynne (2004) are sceptical. Wynne notes sceptically that chimps are slow to interpret the motives of others, and "seem no better able to place themselves imaginatively into the shoes (or paws or hoofs) of another individual than are autistic children" (Wynne 2004, p. 7). However, Rowlands discusses a study which seems to show that one (exceptional) chimp understood that other chimps were concealing food and pretending not to (Rowlands 2007, p.72).

[43] Byrne (1995), Chapter 8. Although the mirror is a vivid, solid example of a reflecting surface, these people must have been familiar with their reflections in water, so perhaps one should be sceptical about their alleged reaction.

For a creature acquainted with mirrors, failure to pass the test must indicate one of the following possibilities: lack of at least primitive self-consciousness; an underlying pathology such as prosopagnosia (face-blindness); or, possibly, lack of self-concern due to depression, boredom, or fatigue. This final possibility assumes that one can separate self-recognition from motivation for spot-rubbing behaviour: the chimp has not only to see something unusual about its appearance, but also want to correct it. There are interesting differences between species in this last respect, as Byrne points out:

> Gorillas are gentle in their investigation and tend to "look but not touch", whereas orang-utans and especially chimpanzees are enthusiastic and often destructive in their actions ... [Unless they have] the extra confidence given by very special treatment, [gorillas may appear] rather disengaged from apparatus such as mirrors put in their cages.

Byrne holds that although reticent with mirrors, great apes also have the potential for mirror self-recognition.[44]

I now move on to defend my interpretation of the implications of the mirror test for self-consciousness.

7.4 Deflationary objection to the mirror test as illustrating primitive self-consciousness

The claim that the mirror test strikingly illustrates a primitive form of self-consciousness is disputed by various writers. There are two main lines of objection. The most common is that passing the test does not show that the chimp is in any sense self-conscious; this objection *deflates* the consequences drawn from the mirror test. The contrary objection *inflates* the consequences drawn from the test, holding that language is not required for full self-consciousness; mirror behaviour illustrates full self-consciousness, and so the Analytic Principle is challenged. This section addresses the first, deflationary objection.

The deflationary line of objection is sceptical that self-consciousness, in any sense, should be attributed to creatures with the cognitive capacities of chimps, and holds that their achievements have been overrated.[45] This line of objection rests on one of two standpoints. The first is extreme scepticism about animal intelligence, that regards the chimp as *hard-wired*

[44] Byrne (1995), pp. 116, 117.

[45] Lowe (2000), pp. 179–81, questions, for instance, whether chimps really engage in practical reasoning.

or programmed to respond to the mirror in the way that it does, or *conditioned* to do so through a process of stimulus and response.[46] Since even the monkey's primitive mirror-behaviour is unlikely to be hard-wired, it is quite implausible to regard the more sophisticated response of the chimp as falling into that category. Monkeys perceive, and learn things about their environments, responding to what they perceive and learn in a way that lifts their responses beyond the category of the hard-wired. If a monkey sees a tiger and flees, its behaviour is probably learned, and not a hard-wired reaction, likewise when it recognises in the mirror a creature of the same species that may be a territorial threat. An anemone, in contrast, does not learn, but merely registers information from its environment, responding to the fish that brush against it by moving its arms away, a behaviour that one cannot describe as involving recognition.

The second standpoint is that *the chimp's behaviour is learned*, yet does not exhibit self-consciousness in any sense. The chimp, it is argued, has simply acquired an understanding of how mirrors work. The following versions of the deflationary objection arise from this assumption:

(1) Because autistic subjects pass the mirror test while lacking a "normal self-concept", it is no test at all.[47]

This objection underestimates the capacities of autistics, whose concept of mind is impaired, but not non-existent, and it assumes attribution of a rich, human concept of self-consciousness to the chimp.[48] To reiterate, proto-self-consciousness in chimps is not the full-blown, core variety that humans exhibit.

(2) The chimp might have an internal set of procedural rules, such as "Do not draw the attention of others to food they haven't noticed", which explain their behaviour without implying self-consciousness.

This objection rests on a questionable conception of inner states, and in particular implies an ability to ascribe conscious states in another of the

---

[46] Dupré shows the implausibility of the latter account in connection with the rather less complex behaviour of domestic cats responding to the familiar sound of food being prepared. In Dupré (2002), p. 226.

[47] Wynne (2004), p. 186. Hauser denies that the mirror test "provides any leverage with respect to self-consciousness", apparently on the (confused) grounds that prosopagnosic patients fail to recognise their mirror-reflection, yet may still be regarded as self-conscious (Hauser 2001, p. 123). Budiansky (1998) is also sceptical, again citing the success of autistic children in passing the mirror test.

[48] For discussion of autism in this connection, see Zahavi (2006), Chapter 7.

same species, for instance to ascribe their not having "noticed" something. This ability is associated with self-consciousness and does not precede it, as we will see; self-consciousness and a concept of mind – that is, a concept of other minds – go together. An implausible complexity of procedures would therefore be required to explain observed primate behaviour without ascribing primitive self-consciousness.

(3) Grasp of mirrored space can be separated from mirror self-recognition; creatures can possess the first without the second.[49] This seems to be the view of developmental psychologists who claim that some mastery of mirrored space precedes mirror self-recognition. Evidence for this claim is that young children who do not pass the mirror test may turn appropriately, when something whizzes past their shoulder in the mirror; and that rhesus monkeys, which fail the mirror test, discriminate snakes and food when presented as mirror reflections, moving away from the snake reflection and approaching the food.[50] Byrne reports that after prolonged exposure to mirrors, monkeys can solve problems by using them: "they can learn to *use* mirrors, for instance to see round a corner and identify another monkey there ... This supports the evidence from cases of deception, that monkeys can work out simple, everyday geometry. But they still systematically fail to understand their *own* reflection in a mirror, and continue to react to it as if it were a stranger".[51]

In fact these considerations do not yield an objection, because it is *mirror self-recognition*, and not *grasp of mirrored space without self-recognition*, that indicates proto-self-consciousness – though it is true that a creature cannot have full grasp of mirrors without mirror self-recognition. So my argument is that understanding mirrors implies mirror self-recognition, and that the latter implies at least primitive self-consciousness. The debate thus involves the following opposition:

(1) Chimps understand mirrors, and, in addition, recognise themselves in them.
(2) They understand mirrors, which involves recognising themselves in them.

---

[49] I owe this objection to Martin Davies.
[50] Hauser (2001), p. 125. His description does not make clear whether they move away from the snake, or from its reflection.
[51] Byrne (1995), p. 113.

I defend the second claim. Despite the evidence offered by developmental psychologists and others, grasp of mirrored space involves the ability to infer the real location of the object whose mirror-image is presented, and so creatures lacking mirror self-recognition have at best only an imperfect understanding of mirrors.[52]

The objection to be resisted, therefore, is that there are creatures with a grasp of mirrored space, who nonetheless lack the capacity for mirrored self-recognition – so these capacities are *non-pathologically split*. The first point to be made in response is that evidence for split capacities is equivocal. The examples cited do not show a full grasp of mirrored space combined with lack of self-recognitional ability, but rather show imperfect grasp of mirrored space, in that the creature has only a partial ability to infer the real location of the object whose mirror-image is presented.[53] I am not here concerned with the question of whether the young child's failure to pass the mirror test is evidence against self-recognition; but the fact that they turn appropriately when something whizzes past their shoulder in the mirror is evidence in favour. The phenomenon is analogous to blindsight, and to amnesiac-remembering as described in Chapter 2: there is evidence that the subject sees, or remembers, and evidence that they do not. I believe that there is no deeper fact of the matter about whether the subject exhibits self-recognition. (This is another instance of "no deeper fact of the matter" arguments.) These subjects do not possess one of two separable capacities, but rather less than one (unified) capacity – just as victims of so-called Multiple Personality Disorder do not have more than one personality, but less than one (integrated) personality.[54]

The relatively sophisticated capacities of self-recognition, and of understanding mirrors – and photographs and pictures of oneself – normally

[52] It could be argued that full grasp of mirrored space involves understanding that it is not real space, and understanding mirrors within mirrors; and that since humans are not good at these things, even they do not possess a complete grasp of mirrored space. The position to be criticised here is analogous to diagnoses of alleged failings of human rationality, when set against some inhuman, utopian standard.

[53] Presumably this ability is rule-based, but the question of whether grasp of rules can be tacit must be pursued on another occasion.

[54] David Spiegel, expert on dissociative disorders for DSM-IV (1994) comments on "widespread misunderstanding of the essential psychopathology in [MPD], which is failure of integration of various aspects of identity, memory, and consciousness. The problem is not having more than one personality; it is having less than one personality" (Hacking 1995, p. 18).

arise together, and constitute the same capacity therefore. A full grasp of mirrors is acquired primarily by being able to recognise oneself in them: by moving one's arm, and seeing that the image's arm moves simultaneously, and so on. Since the monkeys in Byrne's study cannot recognise themselves in mirrored space, they cannot be said to have more than a primitive grasp of it. They exhibit the primitive self-locating capacity possessed by non-self-conscious and non-judging creatures – pigeons, sheep, and upwards to monkeys – discussed in Chapter 6. The chimp, in contrast, exhibits a fuller grasp of mirrors, approaching that of the fully self-conscious creatures who invented the artefact – a grasp involving not just primitive self-location, but also a self-conscious creature's ability to recognise its own appearance. Presumably the child or monkey recognises in the mirror the kind of creature whose image is presented – a child or a monkey – if this really counts as recognition. But a full grasp of mirrored space requires self-recognition in the sense exhibited by the chimp – though as we saw, because the latter cannot express this ability linguistically, its self-consciousness remains of a primitive kind.

It is not just that passing the mirror test is a sophisticated behavioural response, and that creatures that exhibit it tend also to be self-conscious, therefore ability to pass the test is not a mere concomitant of self-consciousness. Rather – when combined with linguistic and other capacities that the chimps lack – it exhibits both self-recognition and self-location in something approaching the rich sense characteristic of full self-consciousness.

Just why mirror self-recognition seems such a striking illustration of self-consciousness still appears to require explanation. My suggestion is that it involves *indirect perception*, analogous to that afforded by photographs, paintings, periscopes, videos, sound recordings, or shadows, and that this requires relatively sophisticated skills. It is only by indirect perception that a creature can *recognise its own appearance*, whether visual or aural. Realising that one is the person whose leaky bag of sugar is causing the trail in the supermarket; or the person being shouted out; or the person whose shoes are squeaking – these are cases of self-identification rather than self-recognition.[55] They do not involve recognising one's appearance, but rather, identifying oneself as the

[55] Elias Canetti's absurdist novel *Auto-da-Fé* describes a bizarre failure of self-identification, showing the central character's self-absorption: "Since [Kien] felt not the slightest desire to notice anyone, he kept his eyes lowered or raised above their heads … Before crossing a busy street … he heard someone shouting loudly at someone else. 'Can you tell me where Mut Strasse is?' There was no

origin or target of some activity or process. Animals routinely respond differentially to white spots on their body that they can see, but this is a less sophisticated ability than mirrored self-recognition, and is not a matter of recognising one's appearance. In fact, as arguments in Ch. 6 suggest, it may not count as self-recognition at all.

Self-recognition and self-consciousness are interdependent because they involve knowing how one appears to others, hence the Strawsonian requirement that to be self-conscious is to be able to view oneself as a person among persons. His salutary and persuasive claim is, in effect, that self-consciousness and possession of a concept of other minds constitute two sides of the same capacity.[56] Self-consciousness is conceptually interdependent with possession of a *concept of mind*, therefore. It is implausible that a chimp could recognise itself in a mirror, yet not exhibit any behaviour that involved attributing intelligence to its co-specifics – such as pretence during food-gathering, which shows that it is able to predict what its co-specifics would do if they had noticed the food, or tracking the gaze of others, like the female being groomed by a younger male. These and other abilities indicate proto-self-consciousness, and a primitive concept of mind. This concept is tested for by the "false-belief task", which very young children, autistic subjects, and monkeys and other animals fail.[57] To say that a chimp has a concept of mind is to say that it recognises its co-specifics as persons in some rudimentary sense, with beliefs and desires of their own – beliefs and desires which may conflict with its own.

---

reply. Kien was surprised: so there were other silent people besides himself to be found in the busy streets … 'Excuse me please, could you perhaps tell me where Mut Strasse is?' … Kien applauded [the silent man] whose character was proof against all chances … Then Kien felt a nasty jolt … a small fat man … was bawling up at him … the man of character, who controlled his tongue even in anger, was Kien himself … ." (Canetti 2005, p. 16).

[56] Strawson's original discussion is found in his critique of solipsism in *Individuals*: "it is a necessary condition of one's ascribing states of consciousness, experiences, to oneself … that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself" (Strawson 1959, p. 99).

[57] For instance, Wimmer and Perner (1983). One should speak of a "concept" rather than a "theory of mind", I believe. Possession of a concept of mind underlies the ability to ascribe propositional attitudes, and to explain and predict behaviour – a capacity that scarcely amounts to possession of a theory. Indeed, describing it as a theory gets the issue precisely the wrong way round. It is the intelligent autistic subject who has a *theory* of mind; lacking the normal subject's intuitive, non-theoretical concept, autistics have to interpret the behaviour of others inferentially. On the "theory of mind" debate, see for instance Zahavi (2006), Chapter 7.

7.5 Inflationary objection to the mirror test as illustrating primitive self-consciousness

Thus far I have been concerned to argue, against sceptics about animal intelligence, that passing the mirror test indicates rudimentary self-consciousness. So I have not attempted to gain a quick victory over critics of the Analytic Principle by denying any sense in which animals can exhibit self-consciousness. Instead – to labour the point, perhaps – I have argued that the chimp's ability of mirror self-recognition indicates only *primitive* or *proto*-self-consciousness, and so is not a counter-example to the Analytic Principle. What the chimp exhibits is a primitive kind of self-consciousness falling short of, and derivative from, the human paradigm. This conclusion conforms with philosophical humanism.

Some writers, however, grant not just primitive, but full self-consciousness to chimps. These writers, at the opposite end of the spectrum to sceptics about animal intelligence, offer an inflationary, anthropomorphic interpretation of the mirror test, according to which it justifies attribution of "I"-thoughts to the chimp. So while the chimp does not manifest linguistic grasp of the first person, its behaviour suggests that it does realise what the monkey does not – that "My forehead has a spot on it".

Such is the intractability of the problem of animal thought, however, that this position may also be found consistent with the Analytic Principle. The chimp is regarded as fully self-conscious, but in having "I"-thoughts does in some sense use the first-person, or at least possess a first person perspective. The remainder of this chapter addresses this inflationary interpretation of the chimp's abilities. To consider it properly one would have to examine not only attribution of the first person, but also attribution of concepts of a forehead or a mark – thus engaging with the vast literature on animal intelligence. Therefore the discussion here only forms the sketch of a later, much longer humanistic treatment.

The debate is between what I will term a liberal or anthropomorphist standpoint, and a restrictive or linguistic one, over whether animals have beliefs, intentions, and other intentional states. In the *anthropomorphist* camp is David Velleman, who seems to regard even cats as self-conscious, though their cognitive powers are no higher than those of monkeys. They have a

> consciousness whose content can be put into words only with the help of the first-person pronoun. [A cat] could never catch a mouse if

> it didn't have thoughts representing the world from its own egocentric perspective, thoughts with English-language equivalents such as "I'm gaining on it" or "I've got it". [In a sense it] has first-personal awareness.[58]

His view seems to be that the cat has literal thoughts, whose content has English-language equivalents.

The *linguistic standpoint* of Dummett and Davidson, in contrast, says that thinking implies literal thought and thus concept-possession, which itself requires language-use.[59] This position underlies many of the deflationary objections to primitive self-consciousness in chimps, criticised in the previous section. It holds that animals do not literally think. Dummett maintains that while humans have conceptual thoughts, animals have "protothoughts" consisting of spatial representations. Davidson argues that "Fido thinks that the cat went up that oak tree" cannot be literally true, "unless we suppose the dog has many general beliefs about trees: that they are growing things, that they need soil and water, that they have leaves or needles, that they burn ... without many [such] general beliefs there would be no reason to identify a belief as a belief about a tree, much less an oak tree".[60]

Philosophical humanism, on the interpretation that I am offering, attempts to undercut the debate between anthropomorphist and linguist. This strategy develops John Dupré's remarks:

> It is commonly supposed that the question whether animals ... have minds is perfectly simple to understand, but very difficult to answer ... [However, the difficulty] is hardly at all to do with lack of evidence, but ... with a lack of clarity about what is really involved in the attribution of mental states [to animals].[61]

---

[58] He continues: "[It] can even have a reflexive awareness ... as when it realizes that the tail it has been catching is its own ... [But it lacks] a conception of a creature that it is. [It] is aware of the mouse that it is chasing, but [not] of there being a creature [that is chasing]. When a cat recognises its own tail, it merely forges a mental association between an object seen to its rear and a locus of sensation or motion at its rear end" (Velleman 2006, pp. 255–6). The paradox here is that Velleman holds – on evidence that is unclear – that the cat has a first-person perspective, while denying that it has a conception of self.

[59] Dummett (1993), Chapters 12 and 13; Davidson (1997), pp. 24–5; (1999), pp. 7–8.

[60] Davidson (1985), p. 475.

[61] Dupré (2002), p. 217.

In particular, I believe, it is too readily assumed that it makes sense to ask "What is the chimp's literal thought?" This question assumes that the creature is a language-user, and it is therefore a bad question. One should neither affirm nor deny that chimps have "I"-thoughts, but instead reject both the anthropomorphist and the linguistic standpoints.

Humanism does not exclude the possibility of animal beliefs, but humanists argue that an animal's behaviour can be understood only by comparing its cognitive powers to human ones.[62] Dupré and Gaita dispute the latter claim on the grounds that the behaviour which, at least in part, fixes the meaning of psychological concepts, comes in both human and non-human forms. Hence they deny that we form the concept of intention first in its application to human behaviour and then apply it to animals. Such concepts, they hold, are formed in responses to animals and to human beings together.[63] The result is what I will term a *hybrid account* of psychological concepts. Dupré and Gaita hold that use of mental concepts commits one at some basic level to the sentience and intelligence of certain non-human creatures.

As it stands, the hybrid account is too anthropomorphic, I believe. But the humanist can accept a qualified version that excludes ascription of higher mental capacities such as self-consciousness and intention. The qualified version thus applies to "Fido feels pain", but not to "Fido is self-conscious". The contrast is recognised by Wittgenstein, who, as we saw earlier, was a humanist in core senses (1) and (2) of section 1. As he suggested, while complex thought requires language, simple thought does not. Higher mammals can possess concepts that are manifested in non-linguistic behaviour.[64] They have intelligence and solve problems, and in that sense they think; it is natural to interpret them as having simple beliefs, but with the possible exception of the most intelligent higher primates, they do not have propositional thoughts, or exhibit judgement.

---

[62] This seems to be Hyman's position, when he writes that "we should philosophize about knowledge with competent speakers in mind, and allow our conception of the cognitive powers of animals to be decided by our epistemology, and not *vice versa*" (Hyman 1999, p. 437). My position is not merely that we *should* philosophise with competent speakers in mind, but that we *have* to.

[63] Gaita (2004), p. 60. This would be one reason why Dupré rejects "exceptionalism" and thus, presumably, "humanism".

[64] Wittgenstein (1958), p. 174; see Kenny (1989), pp. 36–7; and DeGrazia (1994), pp. 154–6.

This plausible position – the *thinking without thoughts* position – was advocated by Norman Malcolm, who argued that while the dog can believe that the cat went up the oak tree, only humans have the thought that it did.[65] Malcolm's view strikes many as deeply paradoxical, however. How can animals have contentless thoughts or beliefs – which are typically regarded as attitudes towards propositions? But reiterate, it was suggested that asking for the literal content of a non-language-user's thought is a bad question. The thinking without thoughts position is a possible consequence of that suggestion. The explanation of the nature of animal belief in this connection is a fascinating issue to which a large literature is devoted, and I do not at present have an account that lays to rest the sceptical response. Here I simply outline a possible defence.

That defence begins by acknowledging that beliefs of higher mammals have components. However, these beliefs are not ascribed on the basis of prior ascription of these components, but rather on the animal's manifesting certain perceptual capacities, forms of behaviour, and facial expressions shared with human beings.[66] "The dog believes that the cat went up that tree" does not imply that it picks out and classifies objects corresponding to the terms "cat" and "tree". Rather, the attribution arises from our understanding of the motivation for the dog's behaviour, based on our everyday, rudimentary knowledge of how dogs live and react to creatures and events in their environment. Discriminations that dogs make are not linguistic, but they are in evolutionary terms at the basis of language and permit the attribution to them of aims and beliefs. (By the same token, animal cries are not linguistic, but are at the basis of language.) We assume, with justification, that dogs have certain perceptual capacities, wants, and dislikes; they chase cats and sheep, and bark at things that threaten or upset them.[67] This everyday knowledge of canine behaviour is beautifully illustrated in Lord Dunsany's novella *Dean Spanley*, where a persuasive tale of a dog's life is recounted by the central character, who the story leads us to believe is a reincarnated canine.[68]

The "thinking without thoughts" position is an attempt to capture this hybrid – linguistic and non-linguistic – status. Animal thought has a certain level of complexity, but insufficient to sustain the intentionality

---

[65] Malcolm (1973).
[66] As Glock (2008) argues.
[67] I owe this line of thought to Glock, op. cit.
[68] Dunsany (2008). This position is supported by Dupré (2002), Chapter 10, esp. pp. 225, 235.

characteristic of linguistically expressed thought. As Davidson remarks, concerning co-referential substitution in belief-attributions to animals, "We have no real idea how to settle, or make sense of, these questions". In humans, substituting co-referential terms may convert true attributions ("X believes that Cicero was Roman") into false ones ("X believes that Tully was Roman"); in animals it can result in absurdity. [69] If the cat goes up the tallest tree in the garden, it would be bizarre to say that the dog believes that it has; he knows that his master is at the door, but it is absurd to suggest that he knows also that the president of the bank is at the door.[70] The dog expects his master when he hears his footsteps, but cannot expect that he will return home next month. These more complex thoughts require language.[71]

Animal thought is also lacking in complexity in the restricted range of propositional attitudes that can be ascribed. A dog can know, believe, or see that p, but it is normally implausible to suggest that it can judge or hope that p, though one might say, when the dog brings the owner its lead, that it hopes to be taken for a walk. One attributes concepts to the dog in the same primitive sense that one attributes beliefs, though the characterisation of concepts, and hence their attribution, is highly contestable; it is not as if we disagree merely on who they should be attributed to.[72]

This is a suggested line of response to an apparently intractable problem. But it is tempting, I concede, to elevate chimps above the level of "thinking without thoughts". Chimps, unlike dogs and other mammals, may be said to exhibit judgement in the absence of language. They learn to use different tools in pursuit of ants and termites, and so, perhaps, they judge their prey, and their tools, to be of one kind rather than another. Their problem-solving behaviour resembles that of humans, and approaches the manifestation of judgement; as Glock comments, they interrupt an activity in order to examine a problem, try out a solution, abandon one tool in favour of another. Their gestures, grimaces, and head-scratching display hesitation, displeasure, and satisfaction.[73] These capacities make it tempting to attribute "I"-thoughts to chimps.

---

[69] Davidson (1984b), p. 163. Attributing theoretical scientific beliefs to a non-scientist might result in absurdity, but of a much lesser order.

[70] See Davidson (1984), p. 163. Lowe argues that the dog associates the sound of its owner's approaching footsteps with the experience of having the lead attached and going for a walk (Lowe 2000, p. 179), but his account is consistent with the animal believing that it will be taken for a walk.

[71] See Hacker (2001), p. 61.

[72] In this I take issue with Glock (2008); see also Dupré (1996), p. 331.

[73] Glock (2008). A more sceptical view is found in Lowe (2000), pp. 179–80.

Although I do not endorse it, arguing instead that chimps exhibit primitive self-consciousness without the entertaining of "I"-thoughts, this inflationary standpoint is not obviously inconsistent with the general terms of the Analytic Principle. In my view, however, full self-consciousness arises with language-use, and in particular the ability to self-refer in the first person. Chimps and other animals may approach full self-consciousness. But when we ascribe proto-self-consciousness to animals, we employ a category that is conceptually parasitic on the richer yet more fundamental human paradigm of self-consciousness.

This volume has advocated a position consistent with, and in some senses flowing from, philosophical humanism, defined in opposition to scientism. Humanism in this sense recognises philosophy's distinctive relation to its history, and while the present monograph is not a work in the history of philosophy, it displays the historical dimension necessary to gain proper understanding of any philosophical debate. New insights have been obtained through continuing engagement with such philosophers as Descartes, Locke, Hume, Reid, and Kant.

It is essential to stress that philosophical humanism is anti-scientistic but not anti-scientific. Humanism celebrates the achievements of science. This volume has drawn on psychological investigation in a humanistic way, which separates philosophy and psychology rather than treating them as lying on a continuum. Although the boundary between conceptual and psychological enquiry is fluid, empirical work in scientific psychology does not directly sustain or disprove philosophical claims. For humanists, the analysis and explanation of self-consciousness are distinct, if related, questions; psychology provides invaluable material for philosophical discussion, which illustrates philosophical claims. That, I hope, is how such material has been used in the present volume.

To argue that empirical research does not provide evidence for philosophical claims, and that scientific psychology should acknowledge the value of conceptual investigation of its own presuppositions, says nothing about the relative importance of the disciplines. It is simply an assertion of the distinction between conceptual and empirical enquiry. In this way, I believe, the distinctive contributions of philosophy and psychology to the study of the mind can each be realised. This volume therefore aims to offer a treatment of the problem of self-consciousness that promotes fruitful cross-fertilisation between philosophy and psychology.

Bibliography

Adams, F. (1991), "Causal Contents", in *Dretske and His Critics*, B. McLaughlin (ed.) (1991) (Oxford: Blackwell).

Adorno, T., and Horkheimer, M. (1987), *Dialectic of Enlightenment* (London: Verso).

Alexander, S. (1920), *Space, Time and Deity: Gifford Lectures 1916–18* (London: Macmillan).

Allison, H. (1983), *Kant's Transcendental Idealism* (New Haven: Yale University Press).

Anscombe, G. E. M. (1963), *Intention* (Oxford: Blackwell).

—— (1974), "Memory, 'Experience' and Causation", in Lewis, H. D. (1974), also in her (1981).

—— (1981a), *Collected Philosophical Papers Vol II: Metaphysics and the Philosophy of Mind* (Oxford: Blackwell).

—— (1981b), "The First Person", in her (1981a).

Armstrong, D. (1973), *Belief, Truth and Knowledge* (Cambridge: University Press).

—— and Malcolm, N. (1984), *Consciousness and Causality* (Oxford: Blackwell).

Asher, M. (1999), *Lawrence: The Uncrowned King of Arabia* (London: Penguin).

Audi, R. (1998), *Epistemology* (London: Routledge).

Austin, J. L. (1962), *Sense and Sensibilia* (Oxford: Oxford University Press).

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

Ayers, M. (1991), *Locke: Epistemology and Ontology* (London: Routledge).

Baddeley (1993), "Working Memory and Conscious Awareness", in Collins, A. F., et al. (eds) (1993), pp. 11–28.

—— (1997), *Human Memory: Theory and Practice* (rev. edn., Hove, UK: Lawrence Erlbaum).

Baier, K. (1970), "Smart on Sensations", in Borst (ed.) (1970), pp. 95–106.

Baker, G., and Morris, K. (1996), *Descartes' Dualism* (London: Routledge).

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

Baldwin, T. (1988), "Phenomenology, Solipsism and Egocentric Thought", *Aristotelian Society Supplementary Volume,* LXII, pp. 27–44.

—— (2007), "The Humanism Debate", in Leiter and Rosen (eds) (2007), pp. 671–710.

—— (ed.) (2003), *The Cambridge History of Philosophy 1870–1945* (Cambridge: Cambridge University Press).

Baumann, P. (1999), "The Scottish Pragmatist? The Dilemma of Common Sense and the Pragmatist Way Out", *Reid Studies*, 2 (2): 47–58.

Barwise, J., and Perry, J. (1981), "Situations and Attitudes", *Journal of Philosophy*, 78: 668–91.

Bayne, T., and Chalmers, D. (2003), "What is the Unity of Consciousness?", in Cleeremans and Frith (ed.) (2003).

Bayne, T., Cleeremans, A. and Wilken, P. (eds) (2009), *Oxford Companion to Consciousness* (Oxford: Oxford University Press).

Beisecker, D. (2002), "Some More Thoughts About Thought and Talk: Davidson and Fellows on Animal Belief", *Philosophy*, 77: 115–24.
Bell, D. (1990), *Husserl* (London: Routledge).
—— (1996), "Solipsism and Subjectivity", *European Journal of Philosophy*, 4: 155–74.
—— (2003), "The Solipsism Debates", in Baldwin (ed.) (2003).
Bermúdez, J. (1994), "The Unity of Apperception in the *Critique of Pure Reason*", *European Journal of Philosophy*, 2: 213–40.
—— (1998), *The Paradox of Self-Consciousness* (Cambridge, MA: Bradford/MIT Press).
—— (2002), "The Sources of Self-Consciousness", *Proceedings of the Aristotelian Society*, 102: 87–108.
—— (2003a), "'I'-Thoughts and Explanation: Reply to Garrett", *Philosophical Quarterly*, 53 (212): 432–6.
—— (2003b), "The Elusiveness Thesis, Immunity to Error through Misidentification and Privileged Access", in Gertler (ed.) (2003), pp. 213–32.
—— (2003c), *Thinking Without Words* (New York: Oxford University Press).
—— (2005a), "Evans and the Sense of 'I'", in Bermúdez (ed.) (2005), pp. 164–94.
—— (ed.) (2005), *Thought, Reference and Experience: Themes from the Philosophy of Gareth Evans* (Oxford: Oxford University Press).
Bermúdez, J., et al. (eds) (1995), *The Body and the Self* (Cambridge, MA: Bradford/MIT Press).
Bilgrami, A. (1992), *Belief and Meaning* (Oxford: Blackwell).
—— (1998), "Self-Knowledge and Resentment", in Wright et al. (eds) (1998).
Bittner, R. (1988), "Understanding a Self-Deceiver", in McLaughlin and Rorty(eds) (1988).
Blackburn, S. (1984), *Spreading The Word* (Oxford: Clarendon).
—— (1997), "Has Kant Refuted Parfit?", in J. Dancy (ed.), *Reading Parfit* (Oxford: Blackwell).
Bleuler, E. (1950), *Dementia Praecox or the Group of Schizophrenias,* trans. J. Zinkin (New York: International Universities Press).
Block, N. (ed.) (1981), *Imagery* (Cambridge, MA: MIT Press).
Borst, C. (ed.) (1970), *The Mind-Brain Identity Theory* (London: Macmillan).
Boucher, J. (2001), "'Lost in a Sea of Time': Time-Parsing and Autism", in Hoerl and McCormack (eds) (2001).
Bradl, J. (ed.) (1989), *The Mind of Donald Davidson* (Amsterdam: Rodopi).
Brandom, R. (1994), *Making It Explicit: Reasoning, Representing and Discursive Commitment* (Cambridge, MA: Harvard University Press).
—— (2000), *Articulating Reasons: An Introduction to Inferentialism* (Cambridge, MA: Harvard University Press).
Brandon, S., Boakes, J., Glaser, D., and Green, R. (1998), "Recovered Memories of Childhood Abuse: Implications for Clinical Practice", *The British Journal of Psychiatry*, 172: 296–307.
Braude, S. (1995), *First Person Plural: Multiple Personality and the Philosophy of Mind* (rev. edn., Lanham, MD: Rowman and Littlefield).
—— (1998), "Commentary on 'False Memory Syndrome and the Authority of Personal Memory-Claims'", *Philosophy, Psychiatry and Psychology*, 4(4).
Brewer, B. (1995), "Bodily Awareness and the Self", in Bermúdez et al. (eds) (1995).

—— (2011), "Realism and Explanation in Perception", in Eilan and Roessler (eds) (2011).
Brinck, I. (1997), *The Indexical "I": The First Person in Thought and Language* (Dordrecht/Boston/London: Kluwer Academic Publishers).
Brook, A. (1994), *Kant and the Mind* (Cambridge: University Press).
—— (1999, accessed 2005), "Circularity or Reduction: What Is the Real Issue?" in "Book Symposium on José Bermúdez *The Paradox of Self-Consciousness*", at http://host.uniroma3.it/progetti/kant/field/bermudezsymp_brook.htm.
Brotton, J. (2012), *A History of the World in Twelve Maps* (London: Allen Lane).
Brown, J. (1989), "Floating a New Idea in Relaxation", *The Independent*, 28 March.
Brun-Rovet, E. (2002), "Reid, Kant and the Philosophy of Mind", *Philosophical Quarterly*, 52 (209): 495–510.
Budd, M. (1989), *Wittgenstein's Philosophy of Psychology* (London: Routledge).
Budiansky, S. (1998), *If A Lion Could Talk: How Animals Think* (London: Weidenfeld & Nicolson).
Burge, T. (1988), "Individualism and Self-Knowledge", *Journal of Philosophy,* 85: 649–65.
Burgos, J., and Ribes-Iñesta, E. (eds) (2008), *The Brain-Behavior Nexus: Conceptual Issues* (Proceedings of the 10th Biannual Symposium on the Science of Behaviour), Guadalajara, Mexico.
Butler, Bishop Joseph (1975), *The Analogy of Religion*, in J. Perry (ed.), *Personal Identity* (Berkeley: University of California Press).
—— (1896), *The Works of Joseph Butler*, W. E. Gladstone (ed.) (Oxford: Clarendon Press).
Butterworth, G. (1995), "An Ecological Perspective on the Origins of Self", in Bermúdez et al. (eds) (1995).
Byers, C. P. (2005), *Personal Identity: Temporality and Continuity* (Ph.D. thesis, University of Sydney).
Byrne, A., and Hilbert, D. (eds) (1997), *Readings on Color: The Philosophy of Color* (Cambridge, MA: MIT Press).
Byrne, R. (1995), *The Thinking Ape* (Oxford: Oxford University Press).
Campbell, J. (1993), "The role of physical objects in spatial thinking", in Eilan et al. (eds) (1993), pp. 65–95.
—— (1994), *Past, Space and Self* (Cambridge, MA: MIT Press).
—— (2001a), "Rationality, Meaning and the Analysis of Delusion", *Philosophy, Psychiatry, & Psychology*, 8 (2/3): 89–100.
—— (2001b), "Memory Demonstratives", in Hoerl (ed.) (2001).
Candlish, S. (1996), "Wittgenstein and the Doctrine of Kinaesthesis", *Australasian Journal of Philosophy*, 74 (4): 581–97.
Canetti, E. (2005), *Auto Da Fe*, London: Harvill.
Carl, W. (1997), "Apperception and Spontaneity", *International Journal of Philosophical Studies*, (5):
Carnap, R. (1967), *The Logical Structure of the World*, trans. R. George (Berkeley and Los Angeles: University of California Press).
Carruthers, P. (1989), *Introducing Persons*, London: Routledge.
—— (2000), *Phenomenal Consciousness: A Naturalistic Theory* (Cambridge: Cambridge University Press).
Cash, W. (2001), "Did A. J. Ayer see God when he 'died'?", *Sunday Telegraph* 25 February, "Review", p. 4.

Cassam, Q. (1989), "Kant and Reductionism", *Review of Metaphysics*, 43: 72–106.
—— (1993), "Inner Sense, Body Sense and Kant's 'Refutation of Idealism'", *European Journal of Philosophy*, 1: 111–27.
—— (1995), "Introspection and Bodily Self-Ascription", in Bermúdez et al. (eds) (1995).
—— (1997a), *Self and World* (Oxford: Clarendon).
—— (1997b), "Subjects and Objects", *Philosophy and Phenomenological Research,* 57 (3): 643–8.
—— (2002), "Representing Bodies", *Ratio,* 15: 315–34.
Cervantes, M. (1972), *Exemplary Stories*, trans. C. A. Jones (Middlesex: Penguin).
Chalmers, D. (1996), *The Conscious Mind* (Oxford: Oxford University Press).
Cheney, D., and Seyfarth, R. (1990), *How Monkeys See The World* (Chicago: Chicago University Press).
Chung, M. (1992), *Social Non-Complicity of Schizophrenia* (Ph.D. thesis, Sheffield University).
Chung, M., Fulford, W., and Graham, G. (eds) (2007), *The Philosophical Understanding of Schizophrenia* (Oxford: Oxford University Press).
Clark, A. (2003), *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence* (Oxford: Oxford University Press).
Clark, P., and Wright, C. (eds) (1988), *Mind, Psychoanalysis and Science* (Oxford: Blackwell).
Cleeremans, A., and Frith, C. (2003), *Unity of Consciousness: Binding, Integration and Dissociation* (New York: Oxford University Press).
Cockburn, D. (1997), *Other Times: Philosophical Perspectives on Past, Present and Future* (Cambridge: Cambridge University Press).
Cockburn, D. (ed.) (1991), *Human Beings* (Cambridge: Cambridge University Press).
Cole, J. (1991), *Pride and a Daily Marathon* (London: Duckworth).
—— and Paillard, J. (1995), "Living without Touch and Peripheral Information about Body Position and Movement: Studies with Deafferented Patients", in Bermúdez et al. (eds) (1995).
Collingwood, R. (1938), *The Principles of Art* (Oxford: Oxford University Press).
Collins, A. (1987), *The Nature of Mental Things* (Indiana: University of Notre Dame Press).
—— (1996), "Moore's Paradox and Epistemic Risk", *Philosophical Quarterly*, 46.
—— (1997), "Personal Identity and the Coherence of Q-Memory", *Philosophical Quarterly*, 47 (186): 73–80.
—— (1999), *Possible Experience: Understanding Kant's Critique of Pure Reason* (Berkeley: University of California Press).
Collins, A. F., et al. (eds) (1993), *Theories of Memory* (Hove, UK: Lawrence Erlbaum).
Copenhaver, B., and Schmitt, C. (1992), *Renaissance Philosophy* (Oxford: Oxford University Press).
Corazza, E. (2004), *Reflecting the Mind: Indexicality and Quasi-Indexicality* (Oxford: Clarendon Press).
Cottingham, J. (ed.) (1992), *The Cambridge Companion to Descartes* (Cambridge: Cambridge University Press).

Coval, S. (1969), "Performatory Self-Reference and the Philosophy of Mind", in J. MacIntosh and S. Coval (eds), *The Business of Reason* (London: Routledge and Kegan Paul).
Craig, E. (1987), *The Mind of God and the Works of Man* (Oxford: Oxford University Press).
Creath, R. (2004), "Quine on the Relevance and Meaning of Analyticity", in Gibson (ed.) (2004), pp. 47–64.
Danto, A. (1965), "Basic Actions", *American Philosophical Quarterly,* 2.
—— (1999a), *The Body/Body Problem* (Berkeley: University of California Press).
—— (1999b), "Basic Actions and Basic Concepts", in his (1999a).
—— (1999c), "Action, Knowledge and Representation", in his (1999a).
—— (1999d), "The Body/Body Problem", in his (1999a).
Darwall, S. (1988), "Self-Deception, Autonomy, and Moral Constitution", in McLaughlin and Rorty (eds) (1988).
Davidson, D. (1980), *Essays on Action and Events* (Oxford: Clarendon Press).
—— (1984a), *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press).
—— (1984b), "Thought And Talk", in Davidson (1984a), pp. 155–70.
—— (1985), "Rational Animals", in E. LePore and B. McLaughlin (eds) *Actions and Events* (Blackwell: Oxford), pp. 473–80.
—— (1987), "Knowing One's Own Mind", in *Proceedings and Addresses of the American Philosophical Association*, 60; also in Cassam (ed.) (1994).
—— (1989), "What Is Present To The Mind", in Bradl (ed.) (1989).
—— (1997) "Seeing Through Language", in J. Preston (ed.), *Thought and Language* (Cambridge: Cambridge University Press), pp. 15–27.
—— (1999) "The Emergence of Thought", *Erkenntnis,* 51: 7–17.
Davis, L. (1979), *Theory of Action* (Englewood Cliffs, NJ: Prentice-Hall).
De Caro, M., and Macarthur, D. (eds) (2004), *Naturalism in Question* (Harvard: Harvard University Press).
De Gaynesford, M. (2003), "Is I Guaranteed to Refer?", *Pacific Philosophical Quarterly,* 84 (2): 109–26.
—— (2006), *I: The Meaning of the First Person Term* (Oxford: Clarendon Press).
Deacon, T. (2012), *Incomplete Nature: How Mind Emerged from Nature* (New York: Norton).
Dear, P. (2001), *Revolutionising the Sciences: European Knowledge and its Ambitions, 1500–1700* (Basingstoke: Palgrave Macmillan).
DeGrazia, D. (1994), "Wittgenstein and the Mental Life of Animals", *History of Philosophy Quarterly,* 11 (1): 121–37.
—— (2005), *Human Identity and Bioethics* (Cambridge: Cambridge University Press).
Dennett, D. (1981), *Brainstorms* (Brighton: Harvester).
—— (1981a), "Are Dreams Experiences?" in his (1981), pp. 129–48.
—— (1981b), "Brain Writing and Mind Reading", in his (1981), pp. 39–52.
Derbyshire, D. (2001), "Dolphins, on Reflection, are Smarter than We Thought", *Daily Telegraph*, May 2.
Descartes, R. (1991), *The Philosophical Writings of Descartes III: The Correspondence*, J. Cottingham et al. (ed.) (Cambridge: Cambridge University Press).
—— (1996), *Descartes: Meditations on First Philosophy* (Cambridge: Cambridge University Press).

Devlin, K. (1991), *Logic and Information* (Cambridge: Cambridge University Press).
Dickie, G. (1974), *Art and the Aesthetic: An Institutional Analysis* (Ithaca, NY: Cornell University Press).
Dilman, I. (1988), "Intentions and the Unconscious", in P. Clark and C. Wright (eds) (1988).
Dretske, F. (1981), *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, Bradford Books).
—— (1995), *Naturalizing the Mind* (Cambridge, MA: Bradford/MIT Press).
Dreyfus, H. (2000), "A Merleau-Pontyian Critique of Husserl's and Searle's Representationalist Accounts of Action", *Proceedings of the Aristotelian Society*.
Dreyfus, H., and Dreyfus, S. (1999), "The Challenge of Merleau-Ponty's Phenomenology of Embodiment for Cognitive Science", in G. Weiss and H. Haber (eds), *Perspective on Embodiment: The Intersections of Nature and Culture* (London: Routledge, 1999).
Dummett, M. (1978), "Truth", in *Truth and Other Enigmas* (London: Duckworth).
—— (1981), *Frege: Philosophy of Language,* 2nd edn., (London: Duckworth).
—— (1993), *Origins of Analytical Philosophy* (London: Duckworth).
Dunsany, Lord, (2008), *Dean Spanley* (London: Harper).
Dupré, J. (2004), "The Miracle of Monism", in De Caro and Macarthur (eds) (2004), pp. 36–59.
Eacott, M. (1998), "False Memories and False Memory Syndrome: What is Memory for?" (Reply to Hamilton 1998b) *Philosophy Psychiatry and Psychology,* 4 (4).
Eilan, N. (2001), "Meaning, Truth and the Self: A Commentary on Campbell and Parnas and Sass", *Philosophy, Psychiatry, & Psychology,* 8 (2/3): 121–32.
Eilan, N., McCarthy, R., and Brewer, W. (eds) (1993), *Spatial Representation: Problems in Philosophy and Psychology* (Oxford: Clarendon Press).
Eilan, N. and Roessler, J. (eds.) (2011), *Perception, Causation, and Objectivity* (Oxford: Oxford University Press).
Evans, G. (1982), *The Varieties of Reference* (Oxford: Clarendon Press).
Frege, G. (1967), "The Thought", trans. A. and M. Quinton, in Strawson (1967).
Fuchs, T. (unpublished), "Psychiatry and Phenomenology Today", presented at The Second R. D. Laing Conference, organised by the Royal Society of Psychiatry, Philosophy Special Interest Group, London 21 October 2000.
Furbank (2003), "Their Noon, Their Midnight…", *New York Review of Books*, 12 June 2003, pp. 68–71.
Gaita, R. (2004), *The Philosopher's Dog* (New York: Random House).
Gallagher, S. (2005), *How the Body Shapes the Mind* (Oxford: Oxford University Press).
Gallois, A. (1996), *The World Without, the Mind Within: An Essay on First-Person Authority* (Cambridge: Cambridge University Press).
Gardner, S. (1999), *Kant and the Critique of Pure Reason* (London: Routledge).
Garrett, B. (2003), "Bermúdez on Self-Consciousness", *Philosophical Quarterly*, 53 (210): 96–101.
Gauthier, D. (1984), "Afterthoughts", in D. MacLean (ed.) (1984).
George, A. (ed.) (1989), *Reflections on Chomsky* (Oxford: Blackwell).
Gertler, B. (ed.) (2003), *Privileged Access: Philosophical Accounts of Self-Knowledge* (Aldershot: Ashgate).

Gibson, J. J. (1962), "Observations on Active Touch", *Psychological Review,* 69: 477–91.
—— (1966), *The Senses Considered as Perceptual Systems* (Boston: Houghton-Mifflin).
—— (1968), "On the Difference between Perception and Proprioception", unpublished "purple peril", http://www.huwi.org/gibson/difference.php.
—— (1972), "Note on The Differences between a Sensory Modality and a Perceptual System", unpublished "purple peril", http://www.huwi.org/gibson/modality.php.
—— (1979), *The Ecological Approach to Visual Perception* (Boston: Houghton-Mifflin).
Gibson, R. (ed.) (2004), *The Cambridge Companion to Quine* (Cambridge: Cambridge University Press).
Gill, C. (ed.) (1989), *The Person and the Human Mind in Ancient Philosophy* (Oxford: Oxford University Press).
Glisky, E., Schacter, D., and Tulving, E. (1986), "Computer-Learning by Memory-Impaired Patients: Acquisition and Retention of Complex Knowledge", *Neuropsychologia,* 24: 313–28.
Glock, H.-J. (1997), "Philosophy, Thought and Language", in J. Preston (ed.), *Thought and Language* (Cambridge: Cambridge University Press), pp. 151–69.
—— (1997a), "Truth without People?", *Philosophy,* 72: 85–104.
—— (2000), "Animals, Thoughts and Concepts", *Synthese*, 119.
—— (2008), "Judgments and concepts in animals", in J. E. Burgos and E. Ribes-Iñesta (eds) (2008), pp. 149–77 (available at https://www.zora.uzh.ch/5676/3/Glock_Guadalajara_2008V.pdf).
Glock, H., and Preston, J. (1995), "Externalism and First-Person Authority", *The Monist,* 78: 515–33.
Goldman, A. (1986), *Epistemology and Cognition* (Cambridge, MA: Harvard University Press).
Graham, G. (1998), "Lukacs and Realism After Marx", *British Journal of Aesthetics*, 38 (2).
Grandy, R., and Warner, R. (eds) (1986), *Philosophical Grounds of Rationality* (Oxford: Clarendon).
Gregory, R. (ed.) (1987), *The Oxford Companion to the Mind* (Oxford: Oxford University Press).
Grice, H. P. (1941), "Personal Identity", *Mind,* 50: 330–50.
—— (1971), "Intention and Uncertainty", *Proceedings of the British Academy*, 57: 263–79.
Grice, P., and Strawson, P. (1956), "In Defense of a Dogma," *Philosophical Review,* 65 (2): 141–58.
Gruneberg, M., and Morris, P. (1992), *Aspects Of Memory 1: The Practical Aspects*, 2nd edn., (London and New York: Routledge).
Grush, R. (2000), "Self, World and Space: The Meaning and Mechanisms of Ego- and Allocentric Spatial Representation", *Brain and Mind*, 1: 59–92.
Guttenplan, S. (1994), *Blackwell Companion to the Philosophy of Mind* (Oxford: Blackwell).
Guyer, P. (2008), *Knowledge, Reason and Taste* (Princeton and Oxford: Princeton University Press).

Hacker, P. M. S. (1997), "Davidson on First-Person Authority", *Philosophical Quarterly*, 47: 285–304.
—— (2001), "Wittgenstein and the Autonomy of Humanistic Understanding", in his *Wittgenstein: Connections and Controversies* (Oxford: Clarendon Press), pp. 34–73.
—— (2002), "Is There Anything It Is Like to Be a Bat?", *Philosophy,* 77: 157–74.
Hacking, I. (1995), *Rewriting the Soul: Multiple Personality and the Sciences of Memory* (Princeton: Princeton University Press).
Hamilton, A. (1987), *The Self and Self-Consciousness* (Ph.D. thesis, St Andrews).
—— (1991), "Anscombian and Cartesian Scepticism", *Philosophical Quarterly*, 41 (162): 39–54.
—— (1995), "A New Look at Personal Identity", *Philosophical Quarterly,* 45 (180): 332–49.
—— (1998a), "Phenomenalism and the Self", in J. Skorupski (ed.) *The Cambridge Companion to Mill* (Cambridge: Cambridge University Press).
—— (1998b), "False Memory Syndrome and the Authority of Personal Memory-Claims", and reply to commentators, *Philosophy Psychiatry and Psychology,* 4(4).
—— (2000), "The Authority of Avowals and the Concept of Belief", *European Journal of Philosophy*, 7 (2), April: 20–39.
—— (2003), "'Scottish Commonsense' about Memory: A Defence of Thomas Reid's Direct Knowledge Account", *Australasian Journal of Philosophy*, 81(2): 229–45.
—— (2007a), *Aesthetics and Music* (London: Continuum).
—— (2007b), "Delusion, Understanding and Belief", in M. Chung, W. Fulford and G. Graham (eds), *The Philosophical Understanding of Schizophrenia* (Oxford: Oxford University Press).
—— (2008), "Intention as Prediction and the Authority of Avowals", *Philosophical Explorations*, 11 (1): 23–37.
—— (2013), *Routledge Philosophical Guidebook to Wittgenstein and "On Certainty"* (London: Routledge).
—— (forthcoming, 2014), *The Claim of Art.*
Hampshire, S. (1959), *Thought and Action* (London: Chatto and Windus).
—— (1963), "Disposition and Memory", *International Journal of Psychoanalysis,* 42: 59–68.
Harley, J., and Woodward, D. (eds) (1987), *The History of Cartography* 1 (Chicago: Chicago University Press).
Harman, G. (1976), "Practical Reasoning", *Review of Metaphysics*, 29.
—— (1986), "Willing and Intending", in R. Grandy and R. Warner (eds) (1986).
Harris, H. (1995), *Identity* (Oxford: Clarendon).
Hauser, M. (2001), *Wild Minds: What Animals Really Think* (London: Penguin).
Heal, J. (2002) "On First-Person Authority", *Proceedings of the Aristotelian Society*, 102: 1–20.
—— (2003), *Mind, Reason, and Imagination: Selected Essays in Philosophy of Mind and Language* (Cambridge: Cambridge University Press).
Heil, J. (1978), "Traces of Things Past", *Philosophy of Science,* 45: 60–72.
Henrich, D. (1989), "The Identity of the Subject in the Transcendental Deduction", in E. Schaper and W. Vossenkuhl (eds) (1989).

—— (1982), "Fichte's Original Insight", in *Contemporary German Philosophy* 1, trans. David Lachterman (University Park, PA: Penn State University Press), pp. 15–53.
Hinton, G. E., McClelland, J. L., and Rumelhart, D. E. (1986), "Distributed Representation", in Rumelhart et al. (eds) (1986).
Hoerl, C. (1999), "Memory, Amnesia and the Past", *Mind & Language,* 14: 227–51.
Hoerl, C., and McCormack, T. (eds) (2001), *Time and Memory: Issues in Philosophy and Psychology* (Oxford: Oxford University Press).
Hornsby, J. (1980), *Actions* (London: Routledge & Kegan Paul).
Humberstone, L. (1997), "Two Types of Circularity", *Philosophy and Phenomenological Research*, 57: 249–80.
Hume, D. (1888), *Treatise of Human Nature*, L. Selby-Bigge (ed.) (Oxford: Oxford University Press).
Hurley, S., and Nudds, M. (eds) (2006), *Rational Animals?* (Oxford: Oxford University Press).
Husserl, E. (1989), *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, Second Book, trans. R. Rojcewicz and A. Schuwer (Dordrecht: Kluwer).
Hylton, P. (2007), *Quine: Arguments of the Philosophers* (London: Routledge).
Hyman, J. (1999), "How Knowledge Works", *Philosophical Quarterly*, 49: 433–51.
Iseminger, G. (2004), *The Aesthetic Function of Art* (Cornell: Cornell University Press).
Israel, J. (2008), *Enlightenment Contested: Philosophy, Modernity, and the Emancipation of Man 1670–1752* (Oxford: Oxford University Press).
Jackson, J. (2003), *De Gaulle* (London: Haus Publishing).
Jacobsen, R. (1996), "Wittgenstein on Self-Knowledge and Self-Expression", in *Philosophical Quarterly*, 46: 12–30.
James, W. (1950), *The Principles of Psychology,* Vol. I (New York: Dover).
Janaway, C. (1999), *Self and World in Schopenhauer's Philosophy* (Oxford: Oxford University Press).
Jaspers, K. (1963), *General psychopathology*, rans. J.Hoenig et al., (Manchester: Manchester University Press).
Kagan, J. (2009), *The Three Cultures: Natural Sciences, Social Sciences and the Humanities in the 21st Century* (Cambridge: Cambridge University Press).
Kant, I. (1929), *Critique of Pure Reason*, trans. N. Kemp Smith (London: Macmillan).
Kavka, G., (1983), "The Toxin Puzzle", *Analysis,* 43: 33–6.
—— (1984), "Deterrent Actions and Retaliatory Actions", in D. MacLean (ed.) (1984).
Keefe, R. (2002), "When Does Circularity Matter?", *Proceedings of the Aristotelian Society*, 102: 275–92.
Kelly, S., (2002), "Merleau-Ponty on the Body", *Ratio,* 15: 376–91; also in M. Proudfoot (ed.), *The Philosophy of Body* (London: Blackwell), 2003, pp. 62–76.
Kenny, A. (1979), in "The First Person", in C. Diamond and J. Teichman (eds), *Intention and Intentionality* (Brighton: Harvester).
—— (1989), *The Metaphysics of Mind* (Oxford: Oxford University Press).
Knight, D. (2009), *Making of Modern Science* (Cambridge: Polity).

Kripke, S. (1975), "A Puzzle About Belief", in Margalit (ed.) (1975).
—— (1982), *Wittgenstein on Rules and Private Language* (Cambridge, MA: Harvard University Press).
Lea, F. A. (1993), *The Tragic Philosopher: Friedrich Nietzsche* (London: Athlone Press).
Lear, J. (1988), *Aristotle: The Desire to Understand* (Cambridge: Cambridge University Press).
Leiter, B., and Rosen, M. (eds) (2007), *The Oxford Handbook of Continental Philosophy* (Oxford: Oxford University Press).
Lennon, K. (1990), *Explaining Human Action* (London: Duckworth).
LePore, E., and McLaughlin, B. (eds) (1985), *Actions and Events* (Oxford: Blackwell).
Lewis, D. (1979), "Attitudes *de dicto* and *de se*", *Philosophical Review,* 88 (4): 513–43.
Lewis, H. D. (1974), *Contemporary British Philosophy*, fourth series (London: Allen and Unwin).
Lloyd, G. (2009), *Disciplines In The Making: Cross-Cultural Perspectives on Elites, Learning and Innovation* (New York: Oxford University Press).
Locke, J. (1975), *An Essay Concerning Human Understanding* (Oxford: Clarendon Press).
Longuenesse, B. (2007), "Kant on the Identity of Persons", in *Aristotelian Society Proceedings*, CVII: 149–67.
Lowe, E. J. (1996), "Substance and Selfhood", in his *Subjects of Experience* (Cambridge: University Press).
—— (2000), *An Introduction to the Philosophy of Mind* (Cambridge: Cambridge University Press).
Luria, A. R. (1972), *The Man with a Shattered World* (New York: Basic Books).
Mach, E. (1897), *The Analysis of Sensations*, reprint (La Salle, Il: Open Court).
McClelland, J. (1995), "Constructive Memory and Memory Distortions: A Parallel-Distributed Processing Approach", in D. Schacter (1995).
Macdonald, G. (ed.) (1979), *Perception and Identity: Essays Presented to A.J. Ayer* (London: Macmillan).
MacDonald, P. (1996), *Descartes, Husserl and Radical Conversion* (Ph. D. thesis, Durham University).
—— (1999), *Descartes and Husserl: The Philosophical Project of Radical Beginnings* (New York: SUNY Press).
—— (2003) *History of the Concept of Mind* (London: Ashgate).
McDowell, J. (1984), "De Re Senses", Philosophical Quarterly, 34 (136): 283–94.
—— (1986), "Singular Thought and Inner Space", in McDowell and Pettit (eds) (1986).
—— (1996), *Mind and World* (Cambridge, MA: Harvard University Press).
—— (1997), "Reductionism and the First Person", in J. Dancy (ed.) *Parfit and His Critics* (Oxford: Blackwell), pp. 230–50.
—— (2001), *Mind, Value and Reality,* new edn., (Harvard: Harvard University Press).
McDowell, J., and Petit, P. (eds) (1986), *Subject, Thought and Context* (Oxford: Oxford University Press).
McGinn, C. (1996), "Born Free", in *The New Republic,* 8 April: 39–41.
MacIntyre, A. (1958), *The Unconscious* (London: Routledge & Kegan Paul).

—— (1984), "The Virtues, the Unity of a Human Life and the Concept of a Tradition", in his *After Virtue* 2nd edn., (Notre Dame, IN: University of Notre Dame Press), pp. 204–25.
McLaughlin, B., and Rorty, A. (eds) (1988), *Perspectives on Self-Deception* (Berkeley: University of California Press).
MacLean, D. (ed.) (1984), *The Security Gamble: Deterrence Dilemmas in the Nuclear Age* (Totowa, NY: Rowman and Allenheld).
Malcolm, N. (1963), *Knowledge and Certainty* (Englewood Cliffs, NJ: Prentice-Hall).
—— (1973), "Thoughtless Brutes", *Proceedings and Addresses of the American Philosophical Society,* 46.
Margalit, A. (1975), *Meaning and Use* (Dordrecht: Reidel).
Martin, C. B., and Deutscher, M. (1966), "Remembering", *Philosophical Review,* 75: 161–96.
Martin, J. (2004), *Myths of Renaissance Individualism* (Basingstoke: Palgrave Macmillan).
Martin, M. (1995), "Bodily Awareness: A Sense of Ownership", in Bermúdez et al. (eds) (1995).
—— (2001), "Out of the Past: Episodic Recall as Retained Acquaintance", in C. Hoerl and McCormack (eds) (2001).
Mayes, A. (1992), "Memory and Aging", in Gruneberg and Morris (eds) (1992), pp. 124–46.
Megone, C. (1998), "Aristotle's Function Argument and the Concept of Mental Illness", *Philosophy, Psychiatry, & Psychology,* 5 (3): 187–201.
Mele, A. (1989), "Intention, Belief and Intentional Action", *American Philosophical Quarterly*, 26.
—— (1992a), "Intentions, Reasons and Beliefs: Morals of the Toxin Puzzle", *Philosophical Studies,* 68: 171–94.
—— (1992b), *Springs of Action: Understanding Intentional Behaviour* (New York: Oxford University Press).
Merleau-Ponty, M. (1962), *Phenomenology of Perception*, trans. C. Smith (London: Routledge and Kegan Paul).
—— (1983), *The Structure of Behaviour*, trans. A. Fisher (Pittsburgh: Dickens University Press).
Mill, J. S. (1979), *Collected Works Volume IX: An Examination of Sir William Hamilton's Philosophy* (Toronto and London: University of Toronto Press and Routledge).
—— (1989), "Notes to James Mill's *Analysis of the Phenomena of the Human Mind*", in *Collected Works,* Vol. XXXI, "Miscellaneous Writings" (Toronto: University Press).
Moore, J. (1999), "Misdisquotation and Substitutivity: When Not to Infer Belief From Assent", *Mind,* 108 (430): 335–65.
Moran, R. (1994), "Interpretation Theory and the First-Person", *Philosophical Quarterly,* 44 (175): 154–73.
—— (2001), *Authority and Estrangement: An Essay On Self-Knowledge* (Princeton: Princeton University Press).
Mortimore, G. (ed.) (1971), *Weakness Of Will* (London: Macmillan).
Mounce, H. (1994), "The Philosophy of the Conditioned", *Philosophical Quarterly,* 44 (175): 174–90.

Moyal-Sharrock, D. (2000), "Words as Deeds: Wittgenstein's 'spontaneous utterances' and the dissolution of the explanatory gap", *Philosophical Psychology,* 13 (3): 355–72.
Nagel, T. (1970), *The Possibility of Altruism* (Oxford: Clarendon).
—— (1979), "What Is It Like To Be A Bat?", in *Mortal Questions* (Cambridge: Cambridge University Press).
—— (1986), *The View From Nowhere* (Oxford: Oxford University Press).
Nietzsche, F. (1989), *Beyond Good and Evil*, trans. Helen Zimmern (London: Prometheus Books).
Noe, A. (2006), *Action in Perception* (Cambridge, MA: MIT/Bradford).
Noonan, H. (1989), *Personal Identity* (London: Routledge).
—— (2010), "The Thinking Animal Problem and Personal Pronoun Revisionism", *Analysis,* 70 (1): 93–8.
Norman, R. (2004), *On Humanism* (London: Routledge).
Nozick, R. (1981), *Philosophical Explanations* (Oxford: Clarendon).
O'Brien, L. (1994), "Anscombe and the Self-Reference Rule", *Analysis*, 54: 277–81.
—— (1995a), "The Problem of Self-Identification", *Proceedings of the Aristotelian Society*, 95: 235–51.
—— (1995b), "Evans on Self-Identification", *Nous,* 29 (2): 232–47.
—— (2007), *Self-Knowing Agents* (Oxford, Clarendon).
O'Sullivan, B. (2006), "The *Euthyphro Argument* (9b–11d)", *The Southern Journal of Philosophy*, 44:4, 657–75.
Olson, E. (1994), "Is Psychology Relevant to Personal Identity?", *Australian Journal of Philosophy*, 72.
—— (1997), *The Human Animal: Personal Identity Without Psychology* (New York: Oxford University Press).
—— (accessed 2009), entry on "Personal Identity" in *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/identity-personal/.
O'Shaughnessy, B. (1980), *The Will: A Dual Aspect Theory* (Cambridge: Cambridge University Press).
—— (1995), "Proprioception and the Body Image", in Bermúdez et al. (eds) (1995).
Owens, D. (2000), *Reason Without Freedom* (London: Routledge).
Pakaluk, M. (2002), "A Defence of Scottish Common Sense", *Philosophical Quarterly,* 52 (209): 564–81.
Parfit, D. (1971), "Personal Identity", *Philosophical Review*, 80: 3–27.
—— (1984), *Reasons and Persons* (Oxford: Clarendon).
—— (1995), "The Unimportance of Identity", in Harris (1995).
Peacocke, C. (1979), *Holistic Explanation: Action, Space, Interpretation* (Oxford: Oxford University Press).
—— (1983), *Sense and Content* (Oxford: Clarendon).
—— (1985), "Intention and Akrasia", in Vermazen and Hintikka (ed.) (1985).
—— (1992), *A Study of Concepts* (Cambridge, MA: MIT Press).
—— (1997), "Colour Concepts and Colour Experience", in Byrne and Hilbert(eds) (1997).
—— (1999), *Being Known* (Oxford: Oxford University Press).
Pears, D. (1980), "Intentions as Judgments", in Van Straaten, Z. (ed.) (1980).
—— (1984), *Motivated Irrationality* (Oxford: Oxford University Press).

—— (1985), "Intention and Belief", in Vermazen and Hintikka (eds) (1985).
—— (1988), *The False Prison: A Study of the Development of Wittgenstein's Philosophy, Volume 2* (Oxford: Clarendon Press).
Perry, J. (1979), "The Problem of the Essential Indexical", *Nous*, 13; also in Perry (1993).
—— (1993), *The Problem of the Essential Indexical and Other Essays* (Oxford: Oxford University Press).
—— (2002), *Identity, Personal Identity and the Self* (Indianapolis: Hackett).
Phillips Griffiths, A. (1962–1963), "On Belief", in *Proceedings of the Aristotelian Society 1962–3*; also in Phillips Griffiths (ed.) (1967)
—— (1967), *Knowledge and Belief* (Oxford: University Press).
Pink, T. (1996), *The Psychology of Freedom* (Cambridge: Cambridge University Press).
Platts, M. (1979), *Ways of Meaning: An Introduction to the Philosophy of Language* (London: Routledge and Kegan Paul).
Pollock, J. (1987), *Contemporary Theories of Knowledge* (London: Hutchinson).
Povinelli, D., and Vonk, J. (2004), "We Don't Need a Microscope to Explore the Chimpanzee's Mind", *Mind & Language*, 19 (1): 1–28.
Pryor, J. (1998), "Immunity to Error through Misidentification", *Philosophical Topics,* 26.
Putnam, H. (1989), *Representation and Hilarity* (Cambridge, MA: MIT Press).
Quine, W. (1953), *From a Logical Point of View* (Harvard: Harvard University Press).
Ramsay, F. (1990), "Facts and Propositions", in *Philosophical Papers,* H. Mellor (ed.) (Cambridge: Cambridge University Press).
Ratcliffe, M. (2008), "Touch and Situatedness", *International Journal of Philosophical Studies*, 16 (3): 299–322.
Reed, E. (1988), *James J. Gibson and the Psychology of Perception* (Yale: Yale University Press).
—— (1997), *From Soul To Mind: The Emergence of Psychology from Erasmus Darwin to William James* (Yale: Yale University Press).
Reid, T. (1967), *Philosophical Works* (Hildesheim: Georg Olms).
—— (2002), *Essays on the Intellectual Powers of Man*, D. Brookes (ed.) (Edinburgh: Edinburgh University Press).
Richardson, L. (2009), *What Is Distinctive about the Senses?* (Ph. D. thesis, Warwick University).
—— (forthcoming, 2013) "Bodily Sensation and Tactile Perception", *Philosophy and Phenomenological Research*.
Ricoeur, P. (1984), *Time and Narrative* 1, trans. K. McLaughlin and D. Pellauer (Chicago: University of Chicago Press).
Robinson, J. A. (1992), "Autobiographical Memory", in Gruneberg and Morris (eds) (1992), pp. 223–51.
Rodel, S. (2007), *Self-Consciousness* (Harvard: Harvard University Press).
Rogers, L., and Kaplan, G. (2004), "All Animals Are Not Equal: The Interface Between Scientific Knowledge and Legislation for Animal Rights", in Sunstein and Nussbaum (eds) (2004).
Roos, A. (2004), "An Objection to Gareth Evans' Account of Self-Identity", *Ratio* (New Series), 17: 207–17.
Rorty, A. (1992), "Descartes on Thinking with the Body", in J. Cottingham (ed.) (1992).

Rorty, R. (1970), "Mind-body Identity, Privacy, and Categories", in Borst (ed.) (1970), pp. 187–213.
Rovane, C. (1987), "The Epistemology of First-Person Reference", *The Journal of Philosophy,* 84: 147–67.
—— (1990), "Branching Self-Consciousness", *Philosophical Review*, 99.
Rowlands, M. (2007), *The Nature of Consciousness* (Cambridge: Cambridge University Press).
—— (2008), *The Philosopher and the Wolf* (London: Granta).
Royal College of Psychiatry (1997), "Reported Recovered Memories of Child Sexual Abuse", *Psychiatric Bulletin,* 21: 663–5.
Rumelhart, D., and Norman, D. (1981), "A Comparison of Models", in G. Hinton and J. Anderson (eds), *Parallel Modes of Associative Memory* (Hillsdale, NJ: Lawrence Erlbaum Associates).
Rumelhart, D., et al. (eds) (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1* (Cambridge: MIT Press).
Rumfitt, I. (1994), "Frege's Theory of Predication: An Elaboration and Defense, with Some New Applications", *The Philosophical Review*, 103 (4).
Rundle, B. (1997), *Mind In Action* (Oxford: Clarendon Press).
Russell, B. (1921), *The Analysis of Mind* (London: George Allen and Unwin).
Ryle, G. (1971), "A Puzzling Element in the Notion of Thinking", in his *Collected Papers, Vol. II: Collected Essays 1929–1968* (London: Hutchinson).
—— (2000), *The Concept of Mind*, Harmondsworth, Middlesex: Penguin.
Sacks, M. (1991), "Personal Identity and Unity of Consciousness", in *The Pursuit of Mind*, R. Tallis and H. Robinson (ed.) (Manchester: Carcanet Press).
Sacks, O. (1985), *The Man Who Mistook His Wife For A Hat* (London: Duckworth).
Sainsbury, M. (2002), *Departing from Frege: Essays in the Philosophy of Language* (London: Routledge).
Schacter, D. (1989), "Memory", in M. Posner (ed.), *Foundations of Cognitive Science* (Cambridge, MA: Bradford/MIT Press), pp. 683–726.
—— (1995), *Memory Distortion: How Minds, Brains and Societies Reconstruct the Past* (Cambridge, MA: Harvard University Press).
Schaper, E., and Vossenkuhl, W. (eds) (1989), *Reading Kant* (Oxford: Blackwell).
Schechtman, M. (1994), "The Same and the Same: Two Views of Psychological Continuity", *American Philosophical Quarterly*, 31 (3): 199–212.
Schier, F. (1986), *Deeper Into Pictures* (Cambridge: Cambridge University Press).
Schlick, M. (1985), *General Theory of Knowledge* (La Salle, IL: Open Court) translation of 2nd edn., 1925.
Schopenhauer, A. (1970) *Der handschriftliche Nachlass*, A. Hübscher (ed.) (Frankfurt am Main: Kramer).
Searle, J. (1984), *Minds, Brains and Science* (London: British Broadcasting Corporation).
Sherrington, C. (1973), *The Integrative Action of the Nervous System* (London: Ayer).
Shoemaker, S. (1963), *Self-Knowledge and Self-Identity* (Ithaca: Cornell University Press).
—— (1968), "Self-Reference and Self-Awareness", *Journal of Philosophy*, 65 (19): 555–67; also in his (1984a).
—— (1984a), *Identity, Cause and Mind* (Cambridge: Cambridge University Press).
—— (1984b), "Persons and Their Pasts", in his (1984a).

—— (1996), "Moore's Paradox and Self-Knowledge", in his *The First-Person Perspective And Other Essays* (Cambridge: Cambridge University Press).
Shoemaker, S., and Swinburne, R. (1984), *Personal Identity* (Oxford: Blackwell).
Sidgwick, H. (1882), "Incoherence of Empirical Philosophy", *Mind,* 7: 530–48.
—— (1905) "The Philosophy of Common Sense", in *The Philosophy of Kant and Other Lectures* (London: Macmillan).
Singer, P. (1993), *Practical Ethics,* 2nd edn., (Cambridge: Cambridge University Press).
Skorupski, J. (1993), *English-Language Philosophy 1750–1945* (Oxford: Oxford University Press).
Smith, B. (1995), "Common Sense", in B. Smith and D. Smith (eds) (1995).
Smith, D. (1995), "Mind and Body", in B. Smith and D. Smith (eds) (1995).
Smith, B., and Smith, D. (eds) (1995), *The Cambridge Companion to Husserl* (Cambridge: Cambridge University Press).
Smith, M. (1994), *The Moral Problem* (Oxford: Blackwell).
Smith, P., and Jones, O. R. (1986), *The Philosophy of Mind: An Introduction* (Cambridge: Cambridge University Press).
Snow, C. P. (1993), *The Two Cultures* (Cambridge: Cambridge University Press).
Snowdon, P. (1989), "Persons, Animals and Ourselves", in Gill (ed.) (1989).
—— (1991), "Personal Identity and Brain Transplants in Human Beings", in Cockburn (ed.) (1991).
—— (1998), "Persons, Animals & Bodies", in Bermúdez (ed.) (1998).
Soames, S. (2003), *Philosophical Analysis in the Twentieth Century 1: The Dawn of Analysis* (Princeton: Princeton University Press).
Southgate, N. (1999), *Personal Identity and Human Animals: A New History and Theory* (Ph.D. thesis, Durham University).
Squires, R. (1969), "Memory Unchained", *Philosophical Review,* 77: 178–97.
Stern, D. (1991), "Models of Memory: Wittgenstein and Cognitive Science", *Philosophical Psychology*, 4 (2).
Stout, G. F. (1930), "In What Way Is Memory-Knowledge Immediate?", *Studies in Philosophy and Psychology* (London: Macmillan).
Strawson, P. (1959), *Individuals: An Essay in Descriptive Metaphysics* (London: Methuen).
—— (1966), *The Bounds of Sense* (London: Methuen).
—— (1967), *Philosophical Logic* (Oxford: Oxford University Press).
—— (1971), "Meaning and Truth", in his *Logico-Linguistic Papers* (London: Methuen).
—— (1992), *Analysis and Metaphysics* (Oxford: Oxford University Press).
Stroud, B. (1977), *Hume* (London: Routledge).
Sunstein, C., and Nussbaum, M. (eds) (2004), *Animal Rights: Current Debates and New Directions* (Oxford: Oxford University Press).
Sutton, J. (1998), *Philosophy and Memory Traces* (Cambridge: Cambridge University Press).
Taylor, C. (1979), "Action As Expression", in C. Diamond and J. Teichman (eds), *Intention and Intentionality* (Brighton: Harvester).
—— (1983), "Dwellers in Egocentric Space", review of *The Varieties of Reference* by Gareth Evans, *Times Literary Supplement*, 11 March: 230.
Thalberg, I. (1971), "Acting Against One's Better Judgment", in G. Mortimore (ed.) (1971).

Todes, S. (2001), *Body and World* (Cambridge, MA: MIT Press).
Tolman, E. (1948), "Cognitive Maps in Rats and Men", *Psychological Review*, 55: 189–208.
Tomasello, M., and Call, J. (2006), "Do Chimpanzees Know What Others See – or Only What They Are Looking at?", in S. Hurley and M. Nudds (eds) (2006), pp. 271–384.
Tulving, E. (1983), *Elements of Episodic Memory* (Oxford: Oxford University Press).
Tye, M. (2004), *Consciousness and Persons: Unity and Identity* (Cambridge, MA: MIT Press).
Uhlig, R. (1998), "Bionic Arm Proves We Have the Technology", *Daily Telegraph*, 3 June, p. 16.
Van Gulick, R. (1995), "How Should We Understand the Relation between Intentionality and Phenomenal Consciousness?" *Philosophical Perspectives*, 9: 271–89.
Van Inwagen, P. (1990), *Material Beings* (Ithaca, NY: Cornell University Press).
—— (2001), "'I am Elizabeth Anscombe' Is Not an Identity Proposition", *Metaphysica*, 2 (1): 5–8.
Van Straaten, Z. (ed.) (1980), *Philosophical Subjects: Essays Presented to P. F. Strawson* (Oxford: Clarendon).
Varzi, A. (2006), "A Note on the Transitivity of Parthood", *Applied Ontology*, 1 (2): 41–6.
Velleman, J. (1989), *Practical Reflection* (Princeton: University Press).
—— (2006), "The Centered Self", in *Self to Self: Selected Essays* (Cambridge: Cambridge University Press).
Vermazen, B., and Hintikka, M. (eds) (1985), *Essays on Davidson: Actions and Events* (Oxford: Clarendon).
Vesey, Godfrey (1977), *Personal identity: A Philosophical Analysis* (Ithaca: Cornell University Press).
Von Wright, G. (1971), *Explanation and Understanding* (London: Routledge and Kegan Paul).
Wedgwood, R. (1995), "Theories of Content and Theories of Motivation", *European Journal of Philosophy*, 3 (3): 273–88.
Wickham, C. (2010), *The Inheritance of Rome: A History of Europe from 400 to 1000* (Harmondsworth, Middlesex: Penguin).
Wiggins, D. (1980), *Sameness and Substance* (Oxford: Basil Blackwell).
—— (1987), *Needs, Values, Truth: Essays in the Philosophy of Value* (Oxford: Clarendon Press).
Williams, B. (1956), "Personal Identity and Individuation", *Proceedings of the Aristotelian Society,* 18; reprinted in Williams (1973).
—— (1973), *Problems of the Self* (Cambridge: Cambridge University Press).
—— (2008), "Philosophy as a Humanistic Discipline", in his *Philosophy as a Humanistic Discipline* (Princeton: Princeton University Press), pp. 180–99.
Wimmer, H., and Perner, J. (1983), "Belief about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception", *Cognition*, 13: 103–28.
Wisdo, D. (1987), "Kierkegaard and Euthyphro," *Philosophy*, 62: 221–6.
Wiseman, R. (2010), *Speaking of Oneself: Self-Consciousness and the First-Person Pronoun* (Ph.D. thesis, York University).

Wittgenstein, L. (1958), *Philosophical Investigations* (Oxford: Blackwell).
—— (1969a), *The Blue and Brown Books* (Oxford: Blackwell).
—— (1969b), *On Certainty* (Oxford: Blackwell).
—— (1975), *Philosophical Remarks* (Oxford: Blackwell).
—— (1979), *Wittgenstein's Lectures, Cambridge 1932–1935*, ed. A. Ambrose (Oxford: Blackwell).
—— (1980), *Remarks on the Philosophy of Psychology*, two vols. (Oxford: Blackwell).
—— (1982), *Last Writings on the Philosophy of Psychology* Vol I, (Oxford: Blackwell).
—— (1988), *Wittgenstein's Lectures on the Philosophy of Psychology 1946–7*, P. T. Geach (ed.) (Chicago: Chicago University Press).
Wollheim, R. (1979), "Memory, Experiential Memory and Personal Identity", in G. Macdonald (ed.) (1979).
Wollheim, R., and Hopkins, J. (eds) (1982), *Philosophical Essays on Freud* (Cambridge: University Press).
Wolterstorff, N. (2001), *Thomas Reid and the Story of Epistemology* (Cambridge: Cambridge University Press).
Woudenberg, R. (1999), "Thomas Reid on Memory", *Journal of the History of Philosophy,* 37: 117–35.
—— (ed.) (2004), *Cambridge Companion to Thomas Reid* (Cambridge: Cambridge University Press).
Wright, C. (1983), *Frege's Conception of Numbers as Objects* (Aberdeen: Aberdeen University Press).
—— (1989), "Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics", in A. George (ed.) (1989).
—— (1992), *Truth and Objectivity* (Cambridge, MA: Harvard University Press).
—— (1998) "Self-Knowledge: The Wittgensteinian Legacy", in Wright et al. (eds) (1998), pp.
—— (2001), *Rails to Infinity: Essays on Themes from Wittgenstein's "Philosophical Investigations"*, (Harvard: Harvard University Press).
Wright, C., Macdonald, C., and Smith, B. (1998), *Knowing Our Own Minds* (Oxford: Oxford University Press).
Wynne, C. (2004), *Do Animals Think?* (Princeton: Princeton University Press).
Young, A. (2000), "Wondrous Strange: The Neuropsychology of Abnormal Beliefs", *Mind and Language,* 15 (1): 47–73.
Zahavi, D. (2006), *Subjectivity and Selfhood: Investigating the First-Person Perspective* (Cambridge, MA: Bradford/MIT Press).

Index

Adorno, Theodor, 213
Analytic Principle, 2, 6, 8–9, 10, 12–13, 16–17, 20, 33, 199–200, 204, 207, 214, 218, 224, 229
Anscombe, Elizabeth, 12, 13, 22–5, 31, 77–8, 112–13, 114, 147, 160–1, 162, 163, 173, 174–5, 176, 178, 184, 186, 196
Aristotle, 147, 201, 206, 210, 211
Armstrong, David Malet, 131, 133
avowals, 19, 38, 39, 40–1, 53, 54, 60, 80, 112, 117–18, 126, 130, 172, 180, 195
Ayer, Alfred Jules "Freddie," 94, 95, 101, 156, 168
Ayers, Michael, 93–5, 140

basic action, 127–8, 130, 133, 138, 141
basic movement, 7, 111, 128, 133, 135, 139–41, 144–7
Bell, David, 167
Bermúdez, José, 12, 14–15, 17, 20, 21, 33, 40, 98, 99, 116, 123, 124, 146, 200, 214, 215
  The Paradox of Self-Consciousness, 123
  thought-language principle, 33
Blackburn, Simon, 87–8
Bleuler, Eugen, 181
blindsight, 131, 221
Brandom, Robert, 88
Brewer, Bill, 88
Brotton, Jerry, 188, 189
Butler, Joseph, 22, 75–6, 83, 84–5, 88–9, 93
Byrne, Richard, 218, 220, 222

Campbell, John, 74, 82, 170–1, 176, 178, 192
Carnap, Rudolf, 93
Cassam, Quassim, 52, 61, 95, 133, 136, 140, 142, 143–6, 147, 149, 150, 152–4, 157–8, 160, 163, 172, 177, 178, 182, 183, 186
Cole, Jonathan, 129
conceptual holism, 6, 8, 10, 14–23, 30–3, 41, 54, 74, 75–6, 83–8, 92–3, 101, 105, 135, 140, 162, 164, 167
contrast argument, 137, 152–60
control model, of action, 110, 127–8
Copenhaver, Brian, 201
Craig, Edward, 168–9

Danto, Arthur, 127, 164
Davidson, Donald, 20, 225, 228
De Gaynesford, Max, 101–4
delusion, 142, 177–83
Descartes, René, 12, 77, 97, 149, 150–1, 152–4, 160, 163, 165, 184, 213, 229
  Cartesian, 50, 51, 54, 77, 94, 97–100, 108, 110, 116, 117, 127–8, 129, 151, 154, 162, 163, 165–6, 167, 176, 183–4, 213
  Meditations, 12, 153
Dickie, George, 86
direct knowledge account, of memory, 48–9, 61, 64
direct knowledge account, of proprioception, 113, 124
distinctness of subjects, 64–5
Distinctness Principle, 13, 14, 34–5, 36, 40, 60
Dretske, Fred, 67–8, 69, 70
  Knowledge and the Flow of Information, 67
Dummett, Michael, 225
Dunsany, Lord
  Dean Spanley, 227

Dupré, John, 205, 208–9, 225, 226

Einstein, Albert, 211
elusiveness thesis, 147–56
Euthyphro, 49, 85–8
Evans, Gareth, 2, 11–13, 36, 40, 42, 52, 62, 63–4, 65, 67–8, 74, 76–9, 80, 81, 82–3, 84, 91, 95–6, 98–100, 110–11, 129, 133, 148, 175–6, 177, 185–8, 190, 192, 194, 197

false memory, 57, 64
Frege, Gottlob, 39–40, 165, 166, 172–5, 184, 185, 188, 193
Furbank, Philip Nicholas, 212

Gaita, Raimond, 226
Gallup, Gordon, 214
Gibson, James Jerome, 105–7, 111–12, 121–4, 157–8, 191
Glock, Hans-Johann, 227, 228
Goethe, Johann Wolfgang von, 46
Goodman, Nelson, 208
Grice, Herbert Paul, 15–16, 102

Hacker, Peter, 155, 197, 203, 228
Hacking, Ian, 44, 57
Harley, John Brian, 189
Hegel, Georg Wilhelm Friedrich, 211
Henrich, Dieter, 97
Hoerl, Christoph, 59
holism, conceptual, 10, 14, 16–18, 19–21, 22, 24, 30, 32, 33, 41, 54, 74, 75–6, 83, 85–6, 87, 88, 92–3, 101, 105, 135, 140, 162, 164, 167, 198
Horkheimer, Max, 213
humanism, philosophical, 9, 198–216, 224–9
Hume, David, 76, 79, 87, 88, 92–6, 101, 149, 151–3, 156, 167, 168, 183, 196, 211, 229
  A Treatise of Human Nature, 212
Husserl, Edmund, 108, 109–11, 127, 128, 133, 147, 164, 165, 167

image theory, of memory, 46, 48, 49, 61, 63
image theory, of proprioception, 108, 112–14
information, personal and impersonal concepts of, 53, 64–71
information garbling, 54, 58

James, William, 37–8, 101, 208
  Principles of Psychology, 37
Jaspers, Karl, 181

Kant, Immanuel, 92–4, 92–101, 151–3, 211, 229
  Critique of Pure Reason, 12, 212
Keefe, Rosanna, 21–2

Leibniz, Gottfried Wilhelm, 93
Lichtenberg, Georg Christoph, 77
Locke, John, 51, 54, 85, 88, 89, 93, 99–100, 108, 109, 136, 137–8, 139–40, 144, 149, 229

McDowell, John, 64, 74, 87–8, 148, 185
McGinn, Colin, 206
Mach, Ernst, 93
  The Analysis of Sensations, 149
MacIntyre, Alasdair, 16–17
Malcolm, Norman, 155, 227
Martin, Michael, 117, 124, 125
Marx, Karl, 211
materialism about self-consciousness, 136, 147, 148, 150, 176
materialism about the body, 7, 107–11, 137, 146
Merleau-Ponty, Maurice, 12, 108, 110–11, 119, 128, 135, 147, 149, 151, 152, 164–7
Mill, John Stuart, 58, 79, 85, 90, 93, 100–1, 211
  Notes to James Mill's *"The Analysis of the Human Mind,"* 85
Moore, George Edward, 19, 179–80, 188

Nagel, Thomas, 155, 186
name-users scenario, 2–3, 10, 27–31
Newton, Isaac, 211
Noonan, Harold W., 90, 91
no-priority account, 10, 85–8, 140

no-reference view, of "I," 163, 172, 174–5, 186, 194–8

O'Brien, Lucy, 23, 81, 170, 175, 176, 185, 196

Parfit, Derek, 52, 60–1, 62–4, 72, 79, 81, 85, 93, 95, 99, 204
Peacocke, Christopher, 23, 88
perceptual model, of proprioception, 108, 111, 113–27
Perry, John, 22–3, 31, 32–3, 39
Plato, 86, 210, 211
Popper, Karl, 208
Putnam, Hilary, 208

q-memory, 5, 37, 41, 54, 60–5, 68, 71–4
q-proprioception, 102, 114, 116, 129–34
Quine, Willard Van Orman, 16, 20, 21, 208
  "Two Dogmas of Empiricism," 15

reductionism about persons, 5, 62, 82, 93–5, 98–100
Reid, Thomas, 22, 48, 51, 54, 75–6, 83, 84–5, 88, 89, 93, 154, 165, 229
Rosetta Stone, 70
Russell, Bertrand, 175, 185, 211
Ryle, Gilbert, 158, 179

Sacks, Oliver, 156
Schmitt, Charles Bernard, 201
Schopenhauer, Arthur, 108–9
self-conscious knowledge account, of bodily identity, 7, 41, 91, 107–8, 111, 134–47, 153, 156–7, 159
self-presentation thesis, 136–7, 143, 147–60, 163, 172, 174, 177
self-reference, 1–3, 6, 9, 10, 12–13, 16, 20, 22–34, 53, 102–3, 172–8, 184–5, 196
Shannon, Claude Elwood, 68
Sherrington, Charles Scott, 111, 121
Shoemaker, Sydney, 36, 85, 149, 156
Strawson, Peter Frederick, 11, 15–16, 17, 95, 99, 103–4, 147–9, 151, 160, 166, 184, 223

theory of mind, 223
Tolman, Edward Chace, 190
trace theory, of memory, 60–1
Tulving, Endel, 45

unitary self, 7–8, 75–6, 81–3, 92–101, 167–71
unity of consciousness, 7–8, 75–6, 81–3, 92–101, 167–71

Velleman, David, 224–5

Wiggins, David, 87
Williams, Bernard, 203–5, 209, 211
Wiseman, Rachael, 197
Wittgenstein, Ludwig, 20, 28, 36, 39, 42, 44, 52, 86, 102–3, 114, 125, 126, 129, 134, 145, 146, 148, 153, 162, 169, 171–4, 179–80, 186–8, 194–8, 207, 212–13, 226
  The Blue Book, 38, 42, 129, 194, 195
  On Certainty, 103, 160, 179, 188, 195
  Philosophical Investigations, 195, 196
Woodward, David, 189
Wright, Crispin, 76, 87, 139, 208