

# *The Mereological Fallacy in Neuroscience*

## 3.1 Mereological Confusions in Cognitive Neuroscience

Leading figures of the first two generations of modern brain-neuroscientists were fundamentally Cartesian. Like Descartes, they distinguished the mind from the brain, and ascribed psychological attributes to the mind. The ascription of such predicates to human beings was, accordingly, derivative — as in Cartesian metaphysics. The third generation of neuroscientists, however, repudiated the dualism of their teachers. In the course of explaining the possession of psychological attributes by human beings, they ascribed such attributes not to the mind but to the brain or parts of the brain.

Neuroscientists assume that the brain has a wide range of cognitive, cogitative, perceptual and volitional capacities. Francis Crick asserts:

What you see is not what is *really* there; it is what your brain *believes* is there. . . . Your brain makes the best interpretation it can according to its previous experience and the limited and ambiguous information provided by your eyes. . . . the brain combines the information provided by the many distinct features of the visual scene (aspects of shape, colour, movement, etc.) and settles on the most plausible interpretation of all these various clues taken together. . . . what the brain has to build up is a many-levelled interpretation of the visual scene. . . . [Filling-in] allows the brain to guess a complete picture from only partial information - a very useful ability.<sup>1</sup>

So the brain *has experiences, believes things, interprets clues on the basis of information* made available to it, and *makes guesses*. Gerald Edelman holds that structures within the brain 'categorize, discriminate, and recombine the various brain activities occurring in different kinds of global mappings', and that the brain 'recursively relates semantic to phonological sequences and then generates syntactic correspondences, not from preexisting rules, but by treating rules developing in memory as objects for conceptual manipulation'.<sup>2</sup> Accordingly, the brain *categorizes*; indeed, it 'categorizes its own activities (particularly its perceptual categorizations)', and *conceptually manipulates rules*. Colin Blakemore argues that

We seem driven to say that such neurons [as respond in a highly specific manner to, e.g., line orientation] have knowledge. They have intelligence, for they are able to estimate the probability of outside events — events that are important to the animal in question. And the brain gains its knowledge by a process analogous to the inductive reasoning of the classical scientific method. Neurons present arguments to the brain based on the specific features that they detect, arguments on which the brain constructs its hypothesis of perception.<sup>3</sup>

So the brain *knows things, reasons inductively, and constructs hypotheses* on the basis of arguments, and its constituent neurons are *intelligent, can estimate*

*probabilities*, and *present arguments*. J. Z. Young shared much the same view. He argued that 'we can regard all seeing as a continual search for the answers to questions posed by the brain. The signals from the retina constitute "messages" conveying these answers. The brain then uses this information to construct a suitable hypothesis about what is there.'<sup>4</sup> Accordingly, the brain *poses questions*, *searches for answers*, and *constructs hypotheses*. Antonio Damasio claims that 'our brains can often decide well, in seconds, or minutes, depending on the time frame we set as appropriate for the goal we want to achieve, and if they can do so, they must do the marvellous job with more than just pure reason',<sup>5</sup> and Benjamin Libet suggests that 'the brain "decides" to initiate or, at least, to prepare to initiate the act before there is any reportable subjective awareness that such a decision has taken place'.<sup>6</sup> So brains *decide*, or at least 'decide', and *initiate voluntary action*.

Psychologists concur. J. P. Frisby contends that 'there must be a symbolic description in the brain of the outside world, a description cast in symbols which stand for the various aspects of the world of which sight makes us aware'.<sup>7</sup> So there are *symbols in the brain*, and the brain *uses*, and presumably *understands*, *symbols*. Richard Gregory conceives of seeing as 'probably the most sophisticated of all the brain's activities: calling upon its stores of memory data; requiring subtle classifications, comparisons and logical decisions for sensory data to become perception'.<sup>8</sup> So the brain *sees*, *makes classifications*, *comparisons*, and *decisions*. And cognitive scientists think likewise. David Marr held that 'our brains must somehow be capable of *representing . . . information . . .* The study of vision must therefore include . . . also an inquiry into the nature of the *internal representations* by which we *capture this information* and *make it available as a basis for decisions* about our thoughts and actions.'<sup>9</sup> And Philip Johnson-Laird suggests that the brain 'has access to a partial model of its own capabilities', and has the 'recursive machinery to embed models within models'; consciousness, he contends, 'is the property of a class of parallel algorithms'.<sup>10</sup>

With such broad consensus on the correct way to think about the functions of the brain and about explaining the causal preconditions for human beings to possess and exercise their natural powers of thought and perception, one is prone to be swept along by enthusiastic announcements — of new fields of knowledge conquered, new mysteries unveiled.<sup>11</sup> But we should take things slowly, and pause for thought. We know what it is for human beings to experience things, to see things, to know or believe things, to make decisions, to interpret equivocal data, to guess and to form hypotheses. We understand what it is for people to reason inductively, to estimate probabilities, to present arguments, to classify and categorize the things they encounter in their experience. We pose questions and search for answers, using a symbolism — namely, our language — in terms of which we represent things. But do we know what it is for *a brain* to see or hear, for *a brain* to have experiences, to know or believe something? Do we have any conception of what it would be for *a brain* to make a decision? Do we grasp what it is for a brain (let alone a neuron) *to reason* (no matter whether inductively or deductively), *to estimate probabilities*, *to present arguments*, *to interpret data* and *to form hypotheses* on the basis of its interpretations? We can observe whether a person sees something or other — we look at his behaviour and ask him questions. But what would it be to observe whether a brain sees something — as opposed to

observing the brain of *a person* who sees something. We recognize when a person asks a question and when another answers it. But do we have any conception of what it would be for a brain to ask a question or answer one? These are all attributes of human beings. Is it a new *discovery* that brains also engage in such human activities? Or is it a linguistic innovation, introduced by neuroscientists, psychologists and cognitive scientists, extending the ordinary use of these psychological expressions for good theoretical reasons? Or, more ominously, is it a conceptual confusion? Might it be the case that there is simply *no such thing* as the brain's thinking or knowing, seeing or hearing, believing or guessing, possessing and using information, constructing hypotheses, etc. — that is, that these forms of words make no sense? But if there is no such thing, why have so many distinguished scientists thought that these phrases, thus employed, do make sense?

The question we are confronting is a philosophical question, not a scientific one. It calls for conceptual clarification, not for experimental investigation. One cannot investigate experimentally whether brains do or do not think, believe, guess, reason, form hypotheses, etc. until one knows what it would be for a brain to do so — that is, until we are clear about the meanings of these phrases and know what (if anything) *counts* as a brain's doing these things and what sort of evidence would support the ascription of such attributes to the brain. (One cannot look for the poles of the Earth until one knows what a pole is - that is, what the expression 'pole' means, and also what counts as finding a pole of the Earth. Otherwise, like Winnie-the-Pooh, one might embark on an expedition to the East Pole.) The moot question is: does it make sense to ascribe such attributes to the brain? Is there any such thing as a brain's thinking, believing, etc.? (Is there any such thing as the East Pole?)

In the *Philosophical Investigations*, Wittgenstein made a profound remark that bears directly on our concerns. "*Only of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious.*"<sup>12</sup> This epitomizes the conclusions we shall reach in our investigation. Stated with his customary terseness, it needs elaboration, and its ramifications need to be elucidated.

The point is not a factual one. It is not a matter of fact that only human beings and what behave like human beings can be said to be the subject of these psychological predicates. If it were, then it might indeed be a discovery, recently made by neuroscientists, that brains too see and hear, think and believe, ask and answer questions, form hypotheses and make guesses on the basis of information. Such a discovery would, to be sure, show that it is not only of a human being and what behaves like a human being that one can say such things. This would be astonishing, and we should want to hear more. We should want to know what the evidence for this remarkable discovery was. But, of course, it is not like this. The ascription of psychological attributes to the brain is not warranted by a neuroscientific discovery that shows that, contrary to our previous convictions, brains do think and reason, just as we do ourselves. The neuroscientists, psychologists and cognitive scientists who adopt these forms of description have not done so as a result of *observations* which show that brains think and reason. Susan Savage-Rambaugh has produced striking evidence to show that bonobo chimpanzees, appropriately trained and taught, can ask and answer questions, can

reason in a rudimentary fashion, give and obey orders, and so on. The evidence lies in their behaviour — in what they do (including how they employ symbols) in their interactions with us. This was indeed very surprising. For no one thought that such capacities could be acquired by apes. But it would be absurd to think that the ascription of cognitive and cogitative attributes to the brain rests on comparable evidence. It would be absurd because we do not even know what would show that the brain has such attributes.

Why, then, was this form of description, and the forms of explanation that are dependent upon it, adopted *without argument or reflection*? We suspect that the answer is: as a result of an unthinking adherence to a mutant form of Cartesianism. It was a characteristic feature of Cartesian dualism to ascribe psychological predicates to the mind, and only derivatively to the human being. Sherrington and his pupils Eccles and Penfield cleaved to a form of dualism in their reflections on the relationship between their neurological discoveries and human perceptual and cognitive capacities. Their successors rejected the dualism — quite rightly. But the predicates which dualists ascribe to the immaterial mind, the third generation of brain neuroscientists applied unreflectively to the brain instead. It was no more than an apparently innocuous corollary of rejecting the two-substance dualism of Cartesianism in neuroscience. These scientists proceeded to explain human perceptual and cognitive capacities and their exercise by reference to the brain's exercise of *its* cognitive and perceptual capacities.

It is our contention that this application of psychological predicates to the brain *makes no sense*. It is not that as a matter of fact brains do not think, hypothesize and decide, see and hear, ask and answer questions; rather, it makes no sense to ascribe such predicates *or their negations* to the brain. The brain neither sees, *nor is it blind* — just as sticks and stones are not awake, *but they are not asleep either*. The brain does not hear, but it is not deaf, any more than trees are deaf. The brain makes no decisions, but neither is it indecisive. Only what *can* decide can be indecisive. So, too, the brain cannot be conscious; only the living creature whose brain it is can be conscious - or unconscious. *The brain is not a logically appropriate subject for psychological predicates*. Only a human being and what *behaves* like one can intelligibly and literally be said to see or be blind, hear or be deaf, ask questions or refrain from asking.

Our point, then, is a conceptual one. It makes no sense to ascribe psychological predicates (or their negations) to the brain, save metaphorically or metonymically. The resultant combination of words does not say something that is false; rather, it says nothing at all, for it lacks sense. Psychological predicates are predicates that apply essentially to the whole living animal, not to its parts. It is not the eye (let alone the brain) that sees, but *we* see *with* our eyes (and we do not see with our brains, although without a brain functioning normally in respect of the visual system, we would not see). So, too, it is not the ear that hears, but the animal whose ear it is. The organs of an animal are parts of the animal, and psychological predicates are ascribable to the whole animal, not to its constituent parts.

Mereology is the logic of part/whole relations. The neuroscientists' mistake of ascribing to the constituent *parts* of an animal attributes that logically apply only to

the *whole* animal we shall call 'the mereological fallacy' in neuroscience.<sup>13</sup> The principle that psychological predicates which apply only to human beings (or other animals) as wholes cannot intelligibly be applied to their parts, such as the brain, we shall call 'the mereological principle' in neuroscience.<sup>14</sup> Human beings, but not their brains, can be said to be thoughtful or thoughtless; animals, but not their brains, let alone the hemispheres of their brains, can be said to see, hear, smell and taste things; people, but not their brains, can be said to make decisions or to be indecisive.

It should be noted that there are many predicates that *can* apply both to a given whole (in particular, a human being) and to its parts, and whose application to the one may be inferred from its application to the other. A man may be sunburnt, and his face may be sunburnt; he may be cold all over, so his hands will be cold too. Similarly, we sometimes extend the application of a predicate from a human being to parts of the human body; for I example, we say that a man gripped the handle, and also that his hand gripped the handle; that he slipped, and that his foot slipped. Here there is nothing logically awry. But psychological predicates apply paradigmatically to the *human being (or animal) as a whole*, and *not* to the body and its parts. There are a few exceptions, such as the application of verbs of sensation like 'to hurt' to parts of the body — for example, 'My hand hurts', 'You are hurting my hand'.<sup>15</sup> But the range of psychological predicates that are our concern — that is, those that have been invoked by neuroscientists, psychologists and cognitive scientists in their endeavours to explain human capacities and their exercise — have no literal application to parts of the body. In particular, they have no intelligible application to the brain.

#### FOOTNOTES

1 F. Crick, *The Astonishing Hypothesis* (Touchstone, London, 1995), pp. 30, 32f., 57.

2 G. Edelman, *Bright Air, Brilliant Fire — On the Matter of the Mind* (Penguin, Harmondsworth, 1994), pp. 109f., 130.

3 C. Blakemore, *Mechanics of the Mind* (Cambridge University Press, Cambridge, 1977), p. 91.

4 J. Z. Young, *Programs of the Brain* (Oxford University Press, Oxford, 1978), p. 119.

5 A. Damasio, *Descartes' Error — Emotion, Reason and the Human Brain* (Papermac, London, 1996), p. 173.

6 B. Libet, 'Unconscious cerebral initiative and the role of conscious will in voluntary action', *Behavioural and Brain Sciences*, 8, (1985), p. 536.

7 J. P. Frisby, *Seeing: Illusion, Brain and Mind* (Oxford University Press, Oxford, 1980), pp. 8f. It is striking here that the misleading philosophical idiom associated with the Cartesian and empiricist traditions, namely talk of the 'outside' world, has been transferred from the mind to the brain. It was misleading because it purported

to contrast an inside 'world of consciousness' with an outside 'world of matter'. But this is confused. The mind is not a kind of place, and what is idiomatically said to be *in* the mind is not thereby spatially located (cp. 'in the story'). Hence too, the world (which is not 'mere matter', but also living beings) is not *spatially* 'outside' the mind. The contrast between what is in the brain and what is outside the brain is, of course, perfectly literal and unobjectionable. What is objectionable is the claim that there are 'symbolic descriptions' in the brain.

8 R. L. Gregory, 'The confounded eye', in R. L. Gregory and E. H. Gombrich (eds), *Illusion in Nature and Art* (Duckworth, London, 1973), p. 50.

9 D. Marr, *Vision, a Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco, 1980), p. 3, our italics.

10 P. N. Johnson-Laird, 'How could consciousness arise from the computations of the brain?', in C. Blakemore and S. Greenfield (eds), *Mindivaves* (Blackwell, Oxford, 1987), p. 257.

11 Susan Greenfield, explaining to her television audiences the achievements of positron emission tomography, announces with wonder that for the first time it is possible *to see thoughts*. Semir Zeki informs the Fellows of the Royal Society that the new millennium belongs to neurobiology, which will, among other things, solve the age-old problems of philosophy (see S. Zeki, 'Splendours and miseries of the brain', *Philosophical Transactions of the Royal Society*, B 354, (1999), p. 2054). We shall discuss this view in §14.4.2.

12 L. Wittgenstein, *Philosophical Investigations*, ed. G. E. M. Anscombe and R. Rhees, tr. G. E. M. Auscombe (Blackwell, Oxford, 1953), §281 (see also §§282-4, 357-61). The thought fundamental to this remark was developed by A. J. P. Kenny, 'The homunculus fallacy' (1971), repr. in his *The Legacy of Wittgenstein* (Blackwell, Oxford, 1984), pp. 125—36. For the detailed interpretation of Wittgenstein's observation, see P. M. S. Hacker, *Wittgenstein: Meaning and Mind, Volume 3 of an Analytical Commentary on the Philosophical Investigations* (Blackwell, Oxford, 1990), Exegesis §§281-4, 357-61, and the essay entitled 'Men, minds and machines', which explores some of the ramifications of Wittgenstein's insight. As is evident from chapter 1, he was anticipated in this by Aristotle (D/1 408<sup>b</sup> 2-15, quoted on p. 15 above).

13 Kenny ('Homunculus fallacy', p. 125) uses the term 'homunculus fallacy' to signify the conceptual mistake in question. Though picturesque, it may, as he admits, be misleading, since the mistake is *not* simply that of ascribing psychological predicates to an imaginary homunculus in the head. In our view, the term 'mereological fallacy' is more apt. It should be noted, however, that the error in question is not merely the fallacy of ascribing to a part predicates that apply only to a whole, but is a special case of this more general confusion. As Kenny points out, the misapplication of a predicate is, strictly speaking, not a fallacy, since it is not a form of invalid reasoning, but it leads to fallacies (*ibid.*, pp. 135f.). To be sure, this mereological confusion is common among psychologists as well as neuroscientists.

14 Comparable mereological principles apply to inanimate objects and some of their properties. From the fact that a car is fast, it does not follow that its carburettor is fast, and from the fact that a clock tells the time accurately, it does not follow that its great wheel tells the time accurately.

15 But note that when my hand hurts, I am in pain, not my hand. And when you hurt my hand, you hurt me. Verbs of sensation (unlike verbs of perception) apply to parts of the body; i.e. our body is sensitive, and its parts may hurt, itch, throb, etc. But the corresponding verb phrases incorporating nominals, e.g. 'have a pain (an itch, a throbbing sensation)' are predicable only of the person, not of his parts (in which the sensation is located).