properly speaking a description. One can indeed turn over (mull over) a description in one's mind – but not by 30°, at constant velocity.

21 J. Locke, *An Essay concerning Human Understanding*, II. iii.1.
22 Marr, *Vision*, 20.
23 Ibid., 21.
24 Ibid.
25 Ibid., 31.
26 Ibid., 31–2.
27 Ibid., 36.
28 Ibid., 3.
29 For a detailed and illuminating account, see J. Hyman, *The Imitation of Nature* (Oxford and New York: Basil Blackwell, 1989), ch. 1.
30 For though it is false that there are any images on or in the pineal gland, it at least makes sense to suppose that there are, whereas (for reasons previously elaborated) it is incoherent to suppose that the brain formulates 3D descriptions in the cortex.

6

THE HOMUNCULUS FALLACY

Anthony Kenny

In the *Philosophical Investigations*, Wittgenstein says: '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'.[1] This dictum is often rejected in practice by psychologists, physiologists and computer experts, when they take predicates whose normal application is to complete human beings or complete animals and apply them to parts of animals, such as brains, or to electrical systems. This is commonly defended as a harmless pedagogical device; I wish to argue that it is a dangerous practice which may lead to conceptual and methodological confusion. I shall call the reckless application of human-being predicates to insufficiently human-like objects the 'homunculus fallacy', since its most naïve form is tantamount to the postulation of a little man within a man to explain human experience and behaviour.

One of the first philosophers to draw attention to the homunculus fallacy was Descartes. In his *Dioptrics*, he describes how 'the objects we look at produce very perfect images in the back of the eyes'.[2] He encourages his readers to convince themselves of this by taking the eye of a newly dead man, replacing with paper or eggshell the enveloping membranes at the back, and placing it inside a shutter so as to let light through it into an otherwise dark room. 'You will see (I dare say with surprise and pleasure) a picture representing in natural perspective all the objects outside.' 'You cannot doubt', he continues,

> that a quite similar picture is produced in a living man's eye, on the lining membrane. . . . Further, the images are not only produced in the back of the eye but also sent on to the brain

> . . . and when it is thus transmitted to the inside of our head, the picture still retains some degree of its resemblance to the objects from which it originates.

But he concludes with a warning. 'We must not think that it is by means of this resemblance that the picture makes us aware of the objects – as though we had another pair of eyes to see it, inside our brain.'[3]

To think of the brain as having eyes and seeing the retinal image would be one way of committing the homunculus fallacy. But in spite of warning us against the fallacy at this point, Descartes himself commits it when he comes to discuss the relationship between the soul and the pineal gland:

> If we see some animal approach us, the light reflected from its body depicts two images of it, one in each of our eyes, and these two images form two others, by means of the optic nerves, in the interior surface of the brain which faces its cavities; then from there, by means of the animal spirits with which its cavities are filled, these images so radiate towards the little gland which is surrounded by these spirits, that the movement which forms each point of one of the images tends towards the same point of the gland towards which tends the movement which forms the point of the other image which represents the same part of this animal. By this means the two images which are in the brain form but one upon the gland, which, acting immediately upon the soul, causes it to see the form of this animal.[4]

To speak of the soul encountering images in the pineal gland is to commit the homunculus fallacy; for *pace* Descartes, a soul is no more a complete human being than a brain is. In itself, there is nothing philosophically incorrect in speaking of images in the brain: Descartes himself is anxious to explain that they are very schematic images and not pictures except in a metaphorical sense:

> No images have to resemble the objects they represent in all respects . . . resemblance in a few features is enough, and very often the perfection of an image depends on its not resembling the object as much as it might. For instance, engravings, which consist merely of a little ink spread over paper, represent to us forests, towns, men and even battles and tempests.[5]

There would be nothing philosophically objectionable in the suggestion that these schematic images might be observed by a brain surgeon investigating the gland. What is misleading is the suggestion that these images are visible to the *soul*, whose perception of them constitutes seeing. What is wrong is that exactly the same sorts of problems arise about Descartes' explanation as about his explicandum. To the Aristotelians who preceded Descartes, seeing necessitated a non-mechanistic phenomenon taking place in the eye. Descartes introduced new mechanisms, but in his system the non-mechanistic event in the eye is replaced by a new non-mechanistic reading of patterns in the pineal gland. The interaction between mind and matter is philosophically as puzzling a few inches behind the eye as it is in the eye itself.

One danger, then, of the homunculus fallacy is that in problems concerning perception and kindred matters it conceals what is left to be explained. In the case of Descartes, we are put on our guard by the quaintness of some of the physiology, so that we have no difficulty in discovering the gaps in his account; but the philosophical hiatus can co-exist with much more sophisticated physiological information.

A contemporary expert on perception, Professor R. L. Gregory, at the beginning of his book *The Eye and the Brain*, echoes Descartes' warning against the homunculus fallacy:

> We are so familiar with seeing, that it takes a leap of imagination to realize that there are problems to be solved. But consider it. We are given tiny distorted upside-down images in the eyes, and we see separate solid objects in surrounding space. From the patterns of stimulation on the retinas we perceive the world of objects, and this is nothing short of a miracle.
>
> The eye is often described as like a camera, but it is the quite uncamera-like features of perception which are most interesting. How is information from the eyes coded into neural terms, into the language of the brain, and reconstituted into experience of surrounding objects? The task of eye and brain is quite different from either a photographic or a television camera converting objects merely into images. There is a temptation, which must be avoided, to say that the eyes produce pictures in the brain. A picture in the brain suggests the need of some kind of internal eye to see it but this would need a further eye to see *its* picture . . . and so on in an endless regress of eyes and pictures. This is absurd.

> What the eyes do is to feed the brain with information coded into neural activity – chains of electrical impulses – which by their code and the patterns of brain activity, represent objects. We may take an analogy from written language; the letters and words on this page have certain meanings, to those who know the language. They affect the reader's brain appropriately, but they are not pictures. When we look at something, the pattern of neural activity represents the objects and to the brain *is* the object. No internal picture is involved.[6]

The warning against the fallacy is excellent; but the fallacy is itself implied in the suggestion that the brain knows a language and that it has an object like the objects of perception. A converse fallacy is committed when it is said that we are given tiny, distorted, upside-down images in the eyes and that we perceive patterns of stimulation on the retina. Here it is not a bogus subject of perception which is being supplied, but a bogus object of perception.

The reader may feel that this is completely unfair criticism. The words I have criticized are taken from the first page of a popular book. What is the harm in personifying parts of the body in order to dramatize scientific information which can be stated in completely neutral metaphor-free language?

Whether dramatization is good pedagogy depends on whether the important events happen on or off stage. The overall psychological problem of perception could be stated as follows: how does a human being cope with the available sensory information, and how does he act on it? Or, in one of Gregory's own formulations, how does information control behaviour? Now this is a problem which would still remain to be solved even if we knew every detail of the process of collection and storage of information; and one crucial aspect of it is the same whether the information is in the world, in the retinas, or in the central nervous system. The problem is this: what is the relation between the presence of information in the technical sense of communication theory and the possession of information in the non-technical sense in which one can acquire information about the world by looking?

For if having information is the same as knowing, then containing information is not the same as having information. An airline schedule contains the information about airline departures; but the airline schedule does not *know* the time of departures of the flights. The illiterate slave on whose shaven scalp the tyrant has tattooed his

state secrets does not *know* the information which his head contains.

A category difference is involved here. To contain information is to be in a certain state, while to know something is to possess a certain capacity. A state (such as being a certain shape or size, or having a certain multiplicity or mathematical structure) is something describable by its internal properties; a capacity (such as the ability to run a four-minute mile or to speak French) is describable only by specification of what would count as the exercise of the capacity. States and capacities are of course connected: in the simplest case there is an obvious connection between being a round peg (state) and being able to fit into a round hole (capacity). But the connections are not always (as in that case) analytic; and many forms of expertise consist in knowing which states go with which capacities (e.g. what types of mushroom are poisonous, which alloys will stand which strains).

Knowledge is not a state but a capacity, and a capacity of a unique kind. The state of containing certain information is no doubt connected with the capacity which is knowledge of a certain fact; but the two are not identical, as the earlier examples show. We may wonder what extra is involved in the knowing that *p* over and above containing the information that *p*. What is knowing a capacity to do, and what counts as an exercise of that capacity? Clearly, there is no simple answer. One cannot specify behaviour typical of knowing as one can specify behaviour typical of anger. One cannot even specify behaviour typical of knowing that *p*, for a given *p*; what behaviour the knowledge that *p* will lead to will depend on what one wants. For instance, knowledge that the window is open will lead to different behaviour in the case of someone who wants it open and in the case of someone who wants it shut. To be sure, the verbal utterance of '*p*' is an activity which is uniquely expressive of the knowledge or belief that *p*; but even so, this does not at all mean that anyone who knows that *p* will ever say that *p*.

There is, then, no simple way of specifying how knowledge gets expressed in behaviour and why some pieces of knowledge do not seem to affect one's behaviour at all. Still, to know is to have the ability to modify one's behaviour in indefinite ways relevant to the pursuit of one's goals. It is because the airline schedule does not have any behaviour to be modified by what is written on it that it does not know what the flight times are.

Let us return from knowing to seeing. Seeing, when not illusory, involves knowing: vision might be defined, crudely, circularly, but

not uninformatively, as the acquisition of knowledge in the visual mode. In the Aristotelian tradition, prior to Descartes, it used to be said that it was not the eye that saw, nor the soul, but the whole organism. This was because the normal way to discover whether an organism sees is not just to study its eyes, but to investigate whether its behaviour is affected by changes of light and colour, etc. Consequently, an explanation of seeing must be an explanation not only of the acquisition and storage of information, but also of what makes the containing of this information into knowledge – i.e., its relation to behaviour.

In his paper 'On How So Little Information Controls So Much Behaviour', Gregory well says:

> Perhaps the most fundamental question in the whole field of experimental psychology is: how far is behaviour controlled by currently available sensory information, and how far by information already stored in the central nervous system?[7]

But in that paper he presents a theory of seeing as selection of internal models without saying how the internal models are related to behaviour. He speaks of a model 'calling up the appropriate muscle power' for lifting a certain weight, and of models 'mediating appropriate behaviour',[8] but he nowhere shows how these metaphors might be turned into literal language. What he really explains is how information of a certain type might reach the brain.

Now let us suppose that his explanation of this proves completely correct. Even so, the crucial problem remains; and what is still to be done is masked for the reader, if not for Gregory himself, by the use of homunculus predicates of the brain and the use of intentional or representational or symbolic predicates of items in the brain. Consider the following passage:

> In general the eye's images are biologically important only in so far as non-optical features can be read from the internal models they select. Images are merely patches of light – which cannot be eaten or be dangerous – but they serve as symbols for selecting internal models, which include the non-visual features vital to survival. It is this reading of object characteristics from images that *is* visual perception.[9]

But even if this mechanism is essential for visual perception, it is not visual perception. Selection of internal models would be possible, as seeing would not, in an isolated optical system incapable of

behaviour. This is not just the ordinary-language point – 'we wouldn't *call* such a thing seeing' – it is a methodological point concerning the nature of the problems to be solved and the reasonableness of extrapolations from acquired results. The illusion that what is described is visual perception is encouraged by the use of language such as 'features can be *read*' and '*symbols* for selecting'.

Later in the same paper Gregory writes:

> On this general view perception is not directly of sensory information but rather of the internal models selected by sensory information. Indeed the current perception is the prevailing set of models.[10]

Clearly, it is inadequate to explain what perception is by saying that it is perception not of X but of Y: if I wonder what *perception* is, how am I helped by being told that it is of Y rather than of X? Gregory senses this: that is why his first statement of this thesis is followed by 'indeed' followed by a statement of an incompatible thesis. Perception cannot both be *of* the models and be the models.

So far my objection to the homunculus model has been that it is pedagogically and methodologically dangerous, as helping to cloak the nature of problems to be solved. But there is a more dangerous effect of the model which alone really deserves the name 'fallacy'.

Let us suppose that we waive our objections to the use of human-being predicates for non-human-beings like brains. Let us allow it to be said that the brain is *P*, where *P* is some predicate whose natural application is to whole human beings. (It may, after all, be used in quotes. It usually is – the first time.) There is still an important temptation to be resisted: the temptation to argue from

> This man is *P*

to This man's brain is *P*

or vice versa. Gregory does not always resist this temptation. At the beginning of the quoted paper he argues that learning or storing particular events is always ontogenetic. Naturally stored information, he says, has two origins: ancestral disasters, and previous experience of the individual stored as 'memory'.[11] To prove that storage of particular events is always ontogenetic, he says:

> What is certain is that information gained phylogenetically is always of the general 'skill' kind. We are not able to recall individual events experienced by our ancestors.[12]

And apropos of learning skills such as tennis and piano playing, he says:

> We may be able to recall the odd particular games or concerts, but as skills it is not individual past events which are stored, but rather appropriate behaviour and strategies.[13]

Here the homunculus fallacy is committed thus: 'X remembers that *p*' is being treated as equivalent to 'X has stored the event that *p*'. The only reason given for saying that information about particular events is not stored phylogenetically is that we cannot recall individual events in our ancestors' lives. But this is to argue from 'This man is not *P*' to 'This man's brain is not *P*' which is fallacious, even if the man's brain's being *P* is a necessary condition for his own being *P*.

In another paper, 'Perceptual Illusions and Brain Models', Gregory considers whether the brain is best regarded as a digital or as an analogue device. He writes:

> It is most implausible to suppose that the brain of a child contains mathematical analyses of physical situations. When a child builds a house of toy bricks, balancing them to make walls and towers, we cannot suppose that the structural problems are solved by employing analytical mathematical techniques, involving concepts such as centre of gravity and coefficient of friction of masses. It is far better to make the lesser claim for children and animals: that they behave appropriately to objects by using analogues of senses object-properties, without involving mathematical analyses of the properties of objects and their interactions. Perceptual learning surely cannot require the learning of mathematics. It is far more plausible to suppose that it involves the building of quite simple analogues of relevant properties of objects: relevant so far as they concern the behaviour of the animal or the child.[14]

Here the homunculus fallacy is committed in the sentence, 'Perceptual learning surely cannot require the learning of mathematics.' It is the child that is doing the perceptual learning; what, if anything, is supposed to be learning mathematics is the child's brain. It is implausible that a child building toy bricks should know advanced mathematics; but from this nothing at all follows about what information is contained in the child's brain.

I conclude that there is good reason to heed the warning of Wittgenstein with which this chapter began. The moral is not that the human-being predicates cannot have their use extended at all, but that they must be extended cautiously and self-consciously, and

that if they are extended one may not argue from the application of such a predicate to a whole human being to the application of the transferred predicate to anything other than the whole human being.

POSTSCRIPT

In a brief postscript, I wish to clarify some of the points made above, and to disown some of the theses attributed to me by some of those who read that part of this chapter. First, I do not accuse Professor Gregory of mistaking mechanistic description for conceptual analysis; nor do I think that either the philosopher's answer or the neurophysiologist's answer to the question 'what is perception?' enjoys a privileged status. Second, I do not object to every extension of the application of a predicate from a sentient whole to its parts. Third, I took up no position on the general question whether conscious activities can be said to *be* (nothing but) the micro-structural processes postulated to explain them. I will expand each of these points, and then briefly restate why I call the homunculus fallacy a fallacy.

I do not think that Gregory is under any illusion that he is doing conceptual analysis. I think he is engaged in constructing, and testing experimentally, hypotheses about the mechanisms necessary to explain the phenomena of visual perception. But conceptual analysis is relevant to what he is doing in two ways. First, analysis of the concept of perception is necessary to delimit what are the phenomena to be explained; second, analysis of the concepts of sight and language show that such things as seeing and decoding cannot be done by brains unless we can attribute to brains certain types of behaviour which we can attribute to whole human beings. To attribute such activities to brains without suggesting how the relevant behaviour might be attributable to brains is, I maintained, to mask empirical problems which remain to be solved.

The moral of my chapter, I said, was not that human-being predicates cannot have their use extended at all, but that their use must be extended cautiously. Consequently, I am unmoved if it is pointed out that hands can grasp and hold: such extensions seem to me well within the bounds of caution. Moreover, my objection was not essentially to predicates of wholes being attached to predicates of parts, but to predicates belonging to human beings being attached to non-human beings. The same fallacy could be committed (though my name for it would not be apt) by the incautious application of

human-being predicates to wholes of which human beings are parts, such as communities and states. Populations, like human beings, grow and shrink; but it would obviously be fallacious to argue that a human being was shrinking because the population he belongs to is shrinking, or that a population is growing because every member of it is growing. And states may have intentions which none of their citizens has.[15]

The question whether perception can be said to be identical with physiological processes seems to me to lack a clear sense, and I do not wish to answer it one way or the other. My complaint against Gregory's identification of visual perception with his postulated selection of internal models was not based on a general thesis that perception cannot be identical with a brain process. Though states and capacities are conceptually different, it need not be misleading to say (e.g.) that a peg's ability to fit into round holes *is* its roundness. In the same way, it may be that there is a physiological process – the acquisition of a physiological state – which can be said to *be* visual perception. But no one can claim to have identified such a process until he has brought out its connection with the types of behaviour which are the criteria for the occurrence of visual perception. And this Gregory has not done.

A fallacy, strictly speaking, is a form of argument which can lead from true premisses to a false conclusion. The inappropriate use of predicates, not being a form of argument, is not strictly a fallacy, as I observed. But it leads to a form of argument, which I claimed to detect in Gregory's articles, which is fallacious in the strict sense of the word: the argument that because a certain human-being predicate attaches to a human being it attaches to his brain, or vice versa. The mere inappropriate use of human-being predicates may be called a fallacy in an extended sense, because it may suggest conclusions which are unjustified – notably the conclusion that more has been explained by a psychological theory than has in fact been explained.

Normally, in an adult human being, the ability to see carries with it the ability to say what is seen, though of course not everything which is actually seen is actually talked about. The use of language to report what is seen, like any use of language, is remarkably free from stimulus control – a point which has been repeatedly made, in general terms, by Chomsky. No account of human perception can approach adequacy unless it includes an explanation of this fact. Consequently, even if we knew every detail

of the physiological processes by which visual information reaches the brain, and every detail of the physiological processes by which the linguistic utterance of visual reports is produced, the problem of the relationship between the input and the output would be completely untouched. This problem is a major part of the problem of the physiological explanation of perception, and its existence is masked by talk of the brain reading features of objects from images and calling up appropriate muscle power.

NOTES

1 L. Wittgenstein, *Philosophical Investigations* (Oxford: Basil Blackwell, 1953), § 281.

2 R. Descartes, 'Dioptrics', in *Philosophical Writings*, trans. and ed. E. Anscombe and P. T. Geach (Edinburgh: Nelson, 1954), 239–56, at 244.

3 Ibid., 245–6.

4 R. Descartes, 'Passions of the Soul', in *The Philosophical Works of Descartes*, I, trans. and ed. E. S. Haldane and G. R. T. Ross (Cambridge: Cambridge University Press, 1931), 329–429, at 348.

5 'Dioptrics', 246.

6 R. L. Gregory, *The Eye and the Brain* (London: McGraw-Hill, 1966), 7.

7 R. L. Gregory, in *Bionics Research Reports* I (April 1968), 1 (this paper was also published in C. H. Waddington (ed.), *Towards a Theoretical Biology*, II (Edinburgh: Edinburgh University Press, 1969), 236–46).

8 Ibid., 8.

9 Ibid., 5.

10 Ibid., 8.

11 Ibid., 1 (his quotes).

12 Ibid.

13 Ibid.

14 R. L. Gregory, in *Proceedings of the Royal Society* B, 171 (1968), 279–96, at 294.

15 Cf. L. Wittgenstein, *Zettel*, ed. G. E. M. Anscombe and G. H. von Wright, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1967), 1–48.