THE RELEVANCE OF WITTGENSTEIN'S PHILOSOPHY OF PSYCHOLOGY TO THE PSYCHOLOGICAL SCIENCES

P. M. S. Hacker

1. *The 'confusion of psychology'*

On the concluding page of what is now called 'Part II' of the *Investigations*, Wittgenstein wrote

> The confusion and barrenness of psychology is not to be explained by calling it a "young science"; its state is not comparable with that of physics, for instance, in its beginnings. (Rather with that of certain branches of mathematics. Set theory.) For in psychology there are experimental methods and *conceptual confusion*. (As in the other case, conceptual confusion and methods of proof.)
>
> The existence of the experimental method makes us think we have the means of solving the problems that trouble us; though problem and method pass one another by. (PI p. 232)

This remark was directed against Wolfgang Köhler's reflections in his book *Gestalt Psychology* (1929), the second chapter of which was entitled 'Psychology as a Young Science'. Köhler had noted that the characteristic feature of the development of physics was the transformation of qualitative observations (e.g. of warmth and cold, or of light intensity) into quantitative measurement by means of instruments. This transformation facilitated the discovery of precise functional laws of physics. Köhler approved of behaviourist psychologists who had replaced introspectionist methods by methodical observations of behaviour, but he criticized them for failing to appreciate that the direct experience of the subject nevertheless remains, for the time being, the raw material of the observational psychologist. The discrimination of qualitative types of behaviour is still indispensable, for however useful pneumographic, galvanographic, or plethysmographic methods of measurement may be, they are no substitute for the observational identification of, e.g. anger, fear or anxiety, as experienced by the subject.

Galileo, in the seventeenth century, was able to handle mechanics in quantitative terms, Köhler explained, because the correlation between direct observation of motion and the results of

quantitative measurements of distance, time, weight, etc. was clear. Psychology, however, has not yet achieved this state. We lack the detailed knowledge of psychological phenomena and of their functional relationships with measurable magnitudes pertaining to neurophysiology which might be the basis of a mature science of psychology. So, Köhler proposed, 'if we wish to imitate the physical sciences, we must not imitate them in their contemporary, most developed form; we must imitate them in their historical youth, when their state of development was comparable to our own at the present time.'[1]

It was to this that Wittgenstein was objecting. The state of psychology is not comparable to that of physics in its infancy. Wittgenstein was not setting his face against experimental psychology. Nor was he objecting to the study of neurological causes of psychological malfunctioning or of the neural processes of normal cognitive, perceptual, affective and volitional functions. On the contrary, it was at least an indirect part of his aim to clear away the conceptual confusions that impede advances in these domains. For a multitude of misconceptions of mental processes, of the faculties of the mind, and of the relationship between the mind, the brain and behaviour are a primary cause of what Wittgenstein saw as the barrenness of psychology. Clarification of the psychological concepts that are deployed in psychological investigations is a prerequisite for posing fruitful questions amenable to experimental methods.

Wittgenstein thought the parallel with physics misleading. The psychologist is not like a physicist studying unobservable particles by examining their effects, as in a Wilson cloud-chamber, and the subject is not like a privileged observer, who can look directly at something that is unobservable by others. The physicist observes the phenomena of motion or of electricity, reports his observations, and then constructs theories to explain the phenomena. The psychologist observes the behaviour of his subjects, including their utterances. But it is mistaken to suppose that the seeing, hearing, thinking, feeling, and willing that are his concern are hidden behind the observable behaviour of the subject. It is equally mistaken to suppose that the subject observes them directly,

[1] W. Köhler, *Gestalt Psychology* (Liveright, New York, 1929), p. 32.

and that his utterances describe his observations. This picture of the inner and the outer, and the correlative conception of introspection and privileged access, are misconceived. It is a confusion to suppose that there are two domains, the physical and the mental, each comparable to the other, each populated with objects, events, states and processes – material ones in the first case and immaterial ones in the second. It is equally erroneous, on the rebound from dualism, to suppose that the mental is really the neural in disguise, let alone to suppose that in the fullness of time, psychology will replace gross qualitative psychological descriptions with quantitative neurological ones.

In short, if psychology is to achieve maturity, what it must do is not emulate the methods of physics, but rid itself of conceptual confusions. Wittgenstein gave us an outline of the conceptual scheme we employ in discourse about our psychological powers and their exercise, and methods for extending this sketch when we need to. This has direct bearing on the psychological sciences. In particular, it can serve to rule out a range of psychological theories that fail to respect this conceptual scheme, while simultaneously invoking it. Wittgenstein's philosophy of psychology provides a firm ground for fundamental criticisms of:

1. Physiological psychology, as elaborated by psychologists such as Wundt, James, Ward, Stout and Titchener, which attempted to correlate the deliverances of introspection with physiological processes.
2. Neuroscientific dualism, as advanced by such neuroscientists as Sherrington, Eccles and Penfield, which postulated mind/brain interaction on the Cartesian model.
3. Behaviourism, as propounded by Watson, Tolman, Skinner or Hull.
4. Brain/body dualism, as adopted by numerous current neuroscientists and psychologists, such as Edelman, Kandel or Crick.
5. Cognitive (representational, computational) psychology, which succeeded behaviourism in the 1960s, to which Chomsky in linguistic theory, and Marr in the theory of vision, also contributed.

Before turning to these specific doctrines, it is necessary to sketch out the main elements of the logical-grammar of psychology that Wittgenstein clarified.

2. *Logico-grammatical elucidations*

A variety of clarifications of the nature of language, language acquisition, and linguistic meaning provide the background for Wittgenstein's philosophy of psychology.

The meaning of an expression, with marginal qualifications, is its use. It is also what is understood by anyone who understands or knows what an expression means. And it is what is given by an explanation of meaning. An explanation of meaning, even a humdrum explanation given by means of a series of examples, is a rule – a standard of correctness – for the use of the explanandum.

An ostensive definition is not a link between word and object, or language and reality, but a rule for the use of a word. A sample in an ostensive definition is not described by the ostensive definition, but is a standard of description. So it belongs to the means of representation. It is a perceptible measure, and when used as a measure, it is juxtaposed for perceptual comparison.

Initial language learning is training, which presupposes a wide range of common innate capacities, imitative propensities, and natural responses to stimuli. To learn a language is to learn new forms of behaviour and action, of social interaction and response.

The use of language is a rule-governed practice, context bound and integrated into human activity. Rules of a language, such as the explanations of meaning that a normal speaker gives or recognizes if given, are common or garden standards of use (unlike rules of calculi). Whether a form of words is an expression of a rule, depends upon its role, not on its form. Rules are guides to behaviour. So they cannot be unknown to their followers. For one cannot be guided by unknown rules, cannot consult unknown rules, cannot appeal to unknown rules in justifying or rectifying one's linguistic (or other) behaviour or in criticizing the behaviour of others as incorrect.

Following a rule is the exercise of a two-way ability. There is no such thing as an agent's following a rule if he lacks the power not to follow the rule. When an agent follows a rule, he complies with the rule intentionally. The rule provides his reason, or part of his reason, for doing what he does. One can V for a reason only if one possesses the two-way power to V. A human being can follow a rule unreflectively or even mechanically, but a machine cannot follow a rule or violate

one, either reflectively or mechanically.

So much by way of background. Now for an array of logico-grammatical remarks on psychological concepts.[2] These are not a part of a theory. They are descriptions of normative (rule-governed) connections between concepts (and between uses of words).

(i) Mastery of psychological predicates is not achieved by grasping a private analogue of a public ostensive definition that invokes a sample. For the putative sample would have to be produced by recollection of what an experience was an experience of (e.g. of pain or of seeing red). But recollecting the content of an experience to function as a defining sample presupposes and cannot explain possession of the concept of the experience. In the absence of a criterion of correct recollection, there would be no difference between remembering right what private sample a predicate means [on analogy with '"white" means *that* ☞ □ colour'] and thinking one remembered right. Hence too, there would be no difference between following a rule (a putative ostensive definition) and thinking one was following a rule. Finally, a memory of an experience cannot function as a sample. It is not a possible object of comparison, since it is not perceptible.

(ii) Psychological predicates typically display first/third-person asymmetry. The characteristic first-person present tense use (an *Aüsserung* or avowal) does not rest on introspection conceived as inner sense; nor does it rest on observation of one's own behaviour. It is groundless. The third-person use, by contrast, rests on what the subject says and does.

(iii) The first-person present tense use is commonly expressive – as 'I'm tired' is an expression of weariness, 'I'll do it' of intention, and 'I'm so pleased' of satisfaction. 'I believe that *p*' is commonly a hesitant assertion, not an assertion of hesitancy; and 'I'd like a drink' is normally an expression of a want, not an autobiographical description. But such sentences can also be used, in appropriate settings, to report.

(iv) An avowal, like expressive non-verbal behaviour, is not inductive evidence for the

---

[2] These are stated here briefly and without argument. For detailed elaboration, see P. M. S. Hacker, *Wittgenstein: Meaning and Mind* (Blackwell, Oxford, 1990).

application of the corresponding attribute in a third-person ascription. For inductive evidence presupposes non-inductive identification of the relata in order to establish the inductive correlation. Here it would presuppose non-inductive identification (coupled with the logical possibillity of misidentification) in the first-person case. But that in turn would presuppose the intelligibility of assignment of meaning by reference to a private ostensive definition. Rather, the non-verbal and the verbal expressive behaviour alike are *criteria*, i.e. constitutive evidence, for appropriate psychological attributions.

(v) The criterial link is looser than entailment – the behaviour is *necessarily good evidence* for the presence of the psychological attribute. It is *defeasible*, but if not defeated, it normally suffices for certainty.

(vi) It *makes sense* to ascribe a psychological attribute to another being, truly *or* falsely, only if it is *possible* for that being to display such behaviour as *would* count as good evidence for the ascription of the psychological attribute, i.e. the appropriate forms of behaviour must be in the creature's behavioural repertoire. Hence the limits of thought and experience are the limits of the possible behavioural expression of thought and experience.

(vii) 'First-person (epistemic) authority' and 'privileged access' are misnomers. In saying 'I have a headache', the speaker enjoys no *authority* regarding an object of knowledge about which he is better informed than others. To have an experience is not to have *access* to anything, but to experience something. The speaker's utterance 'I have a pain' is a criterion for a corresponding third-person ascription, but not because he observes and knows directly what an observer cannot observe and can know only indirectly. If anything, the speaker enjoys a form of verdictive power *to decide*, as when he says 'I want a glass of water'. If a person cannot say what he wants, what he has to do is not *find out*, but examine the desirability characteristics of the options and *decide* what *to* want.

(viii) Contrary to traditional conceptions, 'I know ...' here lacks any genuine epistemic role. 'I know that I am in pain' may be an emphatic or concessive utterance, like 'I am indeed in pain' or

'It really does hurt', but it is not an assertion of knowledge analogous to 'I know that he is in pain'. 'I know what I want (believe, would like)' is an expression of decision not a statement of knowledge, as 'I don't know what I want (believe, would like)' is an expression of indecision, not of ignorance. To be sure, doubt is excluded (e.g. 'I doubt whether I am in pain' or 'I'm not sure whether I have a headache'). But it is excluded neither by the possession of knowledge nor by the presence of certainty, but by grammar.

(ix) Mastery of the first-person use is linked with comprehension of the third-person use (and hence with a grasp of the behavioural criteria that warrant it) via grasp of the fact that the first-person use is a reason for others to respond in such-and -such ways, and hence that the avowals of others are a reason for one to act appropriately. Psychological predicates are Janus-faced. To know their use requires mastery of both first-person use and of third-person criterial ascription.

(x) Third-person ascription is warranted by behavioural criteria, but is not normally *inferred* from behavioural criteria. Normally, we see immediately that another is in pain, upset, tired, angry, sad, cheerful, etc. – we see it *expressed*, *manifest*, in their behaviour, and don't infer it from their 'bare bodily movements'. (Similarly, one may identify something immediately, for example a tree as a magnolia, but if asked how one knows, one would cite the characteristic identifying marks.)

(xi) We see other human beings not as embodied minds or animated bodies, but as living creatures with perceptual, volitional and affective powers informed by reason and acting for reasons, behaving purposively and pursuing goals against a backdrop of social norms and values. We naturally see their behaviour as suffused with intentions and with intentionality, not as 'bare bodily movements'. That we do so is no part of any theory (a 'theory of mind', as some psychologists urge), any more than our psychological vocabulary is part of a theoretical vocabulary (of 'folk psychology' as some folk would have it).

(xii) The subject of psychological predicates is neither the mind nor the brain that a human being has, but the animal as a whole.[3] For talk of the mind is a mere *façon de parler* behind which lie

[3] With the exception of verbs of sensation, which are predicable of a person's body and its parts.

the intellectual faculties and their exercise; and the brain is a mere organ – a part of an animal. Nothing the brain can do would satisfy the criteria for ascribing a psychological predicate to it. Unlike the human being whose brain it is, it does not blush, smile, weep, grimace, laugh, or frown in response to the flow of life. To ascribe psychological attributes to the brain is to commit a mereological fallacy – akin to claiming that it is aeroplane's engines, rather than aeroplanes, that fly, or that it is the great wheel of a clock, rather than the clock as a whole, that keeps time.

These observations provide the backcloth against which we can now display the bearing of Wittgenstein's philosophy of psychology on various schools of empirical psychology.

3. *On introspection and early physiological psychology*

Associationist psychology in the eighteenth and early nineteenth century misunderstood the logical character of avowals and reports of experience, of thinking, believing, wanting and intending. In the wake of the Cartesian and empiricist traditions, these philosopher-psychologists conceived of psychology as the science of consciousness, and of consciousness, on the Cartesian model, as encompassing 'everything which we are aware of as happening within us, in so far as we have awareness of it'.[4] The objects of such awareness Descartes denominated 'thoughts', which included the operations of intellect and will, as well as felt sensation, apparent perception, appetite, and mental imagery. Their occurrence and contents, Descartes supposed, are indubitable. According to Locke, it is 'impossible for anyone to perceive without perceiving that he perceives. When we see, hear, smell, taste, feel, meditate or will anything, we know that we do.'[5] This conception of introspection as inner sense was accepted, with modifications regarding its indubitability and infallibility, by associationist psychologists in the nineteenth century.

In the middle of the nineteenth century psychology began to break free from philosophy, and to establish itself as an autonomous experimental science. The first steps were taken by Ernst Weber

[4] Descartes, *Principles of Philosophy*, Pt. I, section 9.

[5] Locke, *An Essay Concerning Human Understanding*, Bk. II, Ch. xxvii, Sect. 9.

and Gustav Fechner in psycho-physics in mid-century in Leipzig, where the first experimental psychological laboratory was established by Wilhelm Wundt in 1879. It was from Wundt's laboratory that experimental psychology spread to Britain and the USA.

Much of the early work in physiological psychology attempted to relate mental events and processes to neural and cortical processes. For the most part, the conception of the mental that was presupposed was Cartesian and Lockean.[6] In particular, it was taken for granted that introspection is akin to a faculty of sense by the exercise of which each person is able to report (fallibly or infallibly) on mental events, states and processes, which can then be correlated with neural ones. James, a leading figure in introducing experimental psychology into the USA, wrote that 'The word introspection need hardly be defined – it means, of course, the looking into one's own mind and reporting what we there discover. *Everyone agrees that we there discover states of consciousness.*'[7] This is the conception of the mind against which Wittgenstein warred. For it leads to a ramifying network of conceptual confusions concerning the 'inner' and its relationship to the 'outer' that infected psychology, namely:

i. the conception of privileged access to the mind and its contents – as if each of us had access to a private peepshow

ii. the conception of private knowledge of our own minds: that the subject of mental attributes has privileged access to his own mind and epistemic authority regarding it.

iii. the consequent thought that the study of psychology is unavoidably secondhand – that a subject sees what passes in his own mind, and reports it to the psychologist

iv. that neural and cortical states and processes, on the one hand, and any mental attribute, on the other, can be inductively correlated on the basis of introspective reports thus conceived.

Against this Wittgenstein argued that introspection is not a 'looking into oneself'. One can *note* or *attend to* how things are with one, but that is not a form of perceptual observation. One may

---

[6] Helmholtz explicitly acknowledged his debt to Locke

[7] W. James, *Principles of Psychology*, Vol. I, p. 185.

note what crosses one's mind when ..., as one may register the increase or decrease of one's pain in the course of the day, or the waxing or waning of love for another over time. But that is not to perceive anything, it is to pay attention to something. Moreover, it is misleading to characterize this direction of attention as a form of *introspection*. For an introspective person is not simply someone who notes what passes in his mind, but rather a self-conscious person, who frequently reflects on his motives and attitudes, and on their explanation, as well as on his past and his relationships with others (cf. PI §587). Introspection is often a route to self-deception rather than to self-knowledge, and even when it does yield self-knowledge, that is not knowledge of one's aches and pains, passing perceptions, and fleeting thoughts that one may avow, express and otherwise exhibit, or keep to oneself.

What the experimental psychologist can do is correlate physiological events and processes with subjects' behavioural manifestations, avowals, and reports of thought and experience. But it is wrong to suppose that psychological attributes uniformly signify states, events and processes that are in one's mind and which one observes *in foro interno* and reports to the psychologist, who can then correlate them with neural events, states and processes. To understand something is not to be in a mental state, but an ability, and the utterance 'Ah, I understand!' is not a report of the onset of a mental state or process, but a signal of understanding – of the dawning of an ability or flash of insight. Similarly, to think something to be so is to believe or opine, and these are neither acts nor states or processes. One cannot perform an act of believing, otherwise one might ask how long it takes to believe that such-and-such. One cannot be in a mental state of being of such-and-such an opinion, for opining lacks what Wittgenstein called 'genuine duration'. And to believe is no process, otherwise it would make sense to ask whether one has finished believing that such-and-such, and one might answer that one was only half way through. What goes on in one's mind when one is thinking, and which might be correlated with neural events and processes are typically *accompaniments* of thinking – not the thinking itself. Contrary to the stream of consciousness writers, the true nature of thinking is not manifest in what goes on in one's mind, but in what one sincerely says one thinks. To

tell another what one thinks is not to report what went on in one's mind while one was thinking. To mean something by what one said or by a gesture one made is not a mental act of any kind, that might be correlated with a neural event. And to say what one meant is not to report a mental event, but to explain the import of one's utterance or gesture. The general concepts of event, state and process are straight-jackets we impose upon psychological phenomena so that we can aspire to correlate them with neural events, states and processes – but confusions ensue when these categories are misapplied.

These misconceptions of introspection characterized not only associationism and early physiological psychology, but the whole dualist tradition – which persisted among cognitive neuroscientists well into the twentieth century.

4. *On dualism – classical and neuroscientific*

Dualism has dogged European philosophy since its inception. It is perhaps the most natural way to think about ourselves. It informed nascent cognitive neuroscience from its inception in the early modern era (e.g. Jean Fernel, and Descartes himself), and was still prominent in the thought of the greatest of early twentieth-century neuroscientists such as Sherrington, Eccles and Penfield. Sherrington, for example, held that the mind has a body, that the mind is the agent of thought, the source of desire, of knowledge and of value, that it is in liaison with the brain, and that it is identical with the self or the 'I'. Eccles also evolved a latter-day form of Cartesian dualism, arguing that the mind and brain interact in the liaison brain, and that voluntary action consists in a mental act of willing that immediately affects the pyramidal cells in the motor cortex. And Penfield held that the interaction of the mind and brain occurs in what, following Hughlings Jackson, he called 'the highest brain mechanism'.

Classical dualism asserted the logical independence of mind and body, conceived as a pair of substances, one immaterial and the other material. In the psychological sciences, dualism was characteristically committed to two-way interaction (although epiphenomenalism, as advanced by Shadworth Hodgson and T. H. Huxley, envisaged only one-way interaction, and did not conceive of

the mind as a substance). According to Descartes, the body that a mind has is viewed as a mechanism in the workings of which the mind can intervene by acting on the pineal gland (or pineal body, as we would say today). This affects the flow of animal spirits (or neurotransmitters), which Descartes conceived of as consisting of small rapidly moving particles, from the ventricles to the muscles. The attributes ascribable to the body were limited to predicates of motion and rest, mechanical function and malfunction, and somatic appearance.

All psychological attributes were ascribed to the mind – it is the mind, not the body, that suffers pain, enjoys perceptual experiences, thinks, imagines, feels passions, desires, decides and forms intentions. The relationship between psychological acts, activities, states, processes and the behaviour of the body was conceived to be causal. So what we see of other people is bare bodily movement, and it is only by analogy with ourselves or by inference to the best explanation that we know that they are subjects of experience and loci of will.

Wittgenstein spent much time exposing the roots of dualism, evidently thinking that here we have the deepest, most natural, and most tenacious misconceptions. Interestingly, the focus of his criticisms was not that the mind is not a substance of any kind. Nor did he concentrate his criticisms on the doctrine that the essence of the mind is consciousness. His criticisms went deeper. He noted, in Aristotelian spirit, that psychological attributes are attributes of an animal as a whole (PI §281). It is not the mind that is in pain, has a stomach-ache or sore-throat, but the human being. The mind cannot be characterized in terms of its thinking and being conscious, since it is the *human being* who thinks and is conscious. Indeed, it was mistaken to suppose that the mind *has a body* – it is the human being, the person, who has a body; and also has a mind. But to have a mind, and to have a body, is not to stand *in a relation* to anything – it is to have and to exercise a range of powers and to have an array of somatic attributes.[8]

Dualism, according to Wittgenstein, errs in holding the relationship between mind and

[8] For elaboration, see P. M. S. Hacker, *Human Nature – the Categorial Framework* (Blackwell, Oxford, 2007), chapter 9.

behaviour to be causal, i.e. external, non-logical. That mistake carries with it a wide-range of commitments. It means that the dualist is forced to subscribe to the intelligibility of a private language. For if the relationship between the mental and behaviour is external, it could be established in one's own case only by inductive correlation, and in the case of others only by analogy with one's own case. But inductive correlation presupposes the possibility of non-inductive identification, and that presupposes possession of the concept of the mental item that is to be correlated with behaviour. But then the only way such a psychological concept could be possessed is if it is defined by private ostensive definition, the sample for which is provided by memory. And that, as Wittgenstein showed, is incoherent.

The conception of the 'inner' as privately owned and epistemically private is likewise incoherent. The subject of experience is not the owner of his experiences. To have a pain is not to stand in a *relation* to a pain – it is for a part of one's body to hurt. A fortiori, it is not to stand in a relation of private inalienable ownership. Indeed, insofar as it makes sense to speak of two people having the same pain, it is perfectly common. It is a muddle to suppose, as Frege did, that 'you can't have my pain, and I can't have your sympathy'. Pains, in this respect, are more like colours than like pennies. The subject of experience is no more a property of an experience that differentiates it from someone else's experience, than the chair is a differentiating property of the colour it has.

The dualist conception of the mental ascribed to subjects of experience *knowledge* of their own experiences and thoughts. But this, as we have noted, is to confuse the grammatical exclusion of ignorance with the empirical presence of knowledge. Classical dualists held that such subjective knowledge was certain (and foundational). But this, Wittgenstein argued, is to confuse the grammatical exclusion of doubt with the empirical presence of certainty.

Because dualism conceived the relationship between the mental and the behavioural to be external, causal, and established by inductive correlation and analogy, it misconstrued the character of human behaviour. Human behaviour is not bare bodily movement caused by the mind's affecting the flow of animal spirits or neurotransmitters. It is animate behaviour, not the mechanical effects of

an embodied anima. We see human behaviour, mien and expression as *informed* by, not as *caused* by, thought, feeling, purpose and intention, in the context of complex social conventions. That we so see it is not a matter of inference, but an aspect of the human form of life.

Since dualism misconstrued the relationship between the mental and behaviour, the standard psychological accounts of voluntary action are, Wittgenstein noted, misguided. He criticized both the ideo-motor explanation of voluntary action that was advanced by James (and Russell), and the innervationist modification of the ideo-motor theory advanced by Wundt, Helmholtz and Mach, who held that although images of kinaesthetic sensations are necessary for voluntary movement, they need to be supplemented by feelings of innervation or directed impulse.[9] But what distinguishes a voluntary movement from an involuntary one is not the presence of an idea of a kinaesthetic sensation that causes the movement, any more than it is the presence of an act of will. To make a movement is not to make one's limb or body move. A voluntary movement (an action) is no more an involuntary one *plus* something (an idea of a kinaesthetic sensation, an act of will) than perceiving something is a hallucination plus something (e.g. an object that coincides with, and causes, the content of the sensory experience).

5. *Brain/body dualism – or dualism with a materialist face*

The next generation of cognitive neuroscientists, such as Blakemore, Crick, Damasio, Edelman, Kandel, Koch and Young, rejected the Cartesian dualism of their predecessors. They repudiated any supposition that the mind is an immaterial substance in causal interaction with the brain. However, in their endeavours to explain human perceptual, cogitative and volitional powers, they attributed a wide range of psychological predicates to the brain. The brain is variously said to know, believe, think, remember, hypothesize, reason, pose questions, search for answers, and decide. Psychologists, e.g. Gregory or Frisby, concur: they contend there are symbols in the brain, which the brain uses to

[9] For detailed discussion of Wittgenstein's objections to such causal accounts of voluntary behaviour, see P. M. S. Hacker, *Wittgenstein – Mind and Will* (Blackwell, Oxford, 1996), Essay VII.

represent the 'external world', and in perception, the brain classifies, compares, makes logical decisions and constructs hypotheses. And computational cognitive scientists, like Marr, went along with this, claiming that the brain uses internal representations, processes information, operates on symbolic representations and produces descriptions.

This scientific conception is a *mutation* of Cartesianism. Although these scientists reject the Cartesian conception of the mind as an immaterial substance, they retain intact the main *structural* features of Cartesian psychology. All they do is replace immaterial substance with grey glutinous matter, leaving everything else the same – in effect adopting what one might ironically call 'brain/body dualism'. For a wide array of psychological attributes which Cartesian dualists ascribe to the mind, brain/body dualists ascribe to the brain. The Cartesian dualists misguidedly viewed the relationship between the mind and the body, as well as between the mind and behaviour, as external or causal. Their materialist successors, perfectly correctly, view the relationship between the brain and the rest of the body, and between the brain and behaviour, as external or causal. But they erroneously treat the brain as a *res cogitans*. Hence they conceive of perception as the causation of internal representations *in the brain*, of self-consciousness as a self-scanning power *of the brain*, of voluntary movements as movements caused *by the brain's decisions*, and of human behaviour as bare bodily movement. Unsurprisingly, they have difficulty in accounting for knowledge of 'the external world', knowledge of 'other minds', and consciousness itself.

It is not difficult to reconstruct the kind of criticisms that Wittgenstein might have advanced. In the first place, brain/body dualism commits a mereological fallacy. Where Descartes ascribed psychological attributes to the mind, crypto-Cartesian scientists ascribe much the same functions to the brain – which is but a part of a human being. Moreover, both do so in order to *explain* the psychological functions of human beings. But not only is it mistaken to ascribe such attributes to the brain, it fails to explain anything. For if one tries to explain how human beings see by reference to the mind's seeing images on the pineal gland (as Descartes did) or by reference to the brain's apprehending the image it creates (as contemporary scientists such as Kandel, Schwartz or Crick do)

the task of explaining how the mind or brain can see or apprehend replaces that of explaining how the human being sees. But, quite apart from this regress, it makes no sense to ascribe psychological predicates to anything less than the animal as a whole. It is *the sentient creature*, not its brain, that has sensations; sees or is blind; hears or is deaf; is conscious or unconscious. A living brain is not a limiting case of a mutilated human being, and a dead brain is not a human corpse.

Wittgenstein would evidently have criticized representational theories of perception of the kind that psychologists and neuroscientists are explicitly or implicitly committed to. What is perceived is not a representation of anything, save in cases in which one perceives a picture or the like. The object of perception is neither in the mind nor in the brain, but in the environment. It doubtless causes a variety of neural events in our sense organs and brain, but these are not what we perceive. Rather, they are what make it possible for us to perceive what we perceive. To perceive something is not, *pace* Helmholtz, the conclusion of an unconscious inference. It is not, *pace* Gregory, to construct a hypothesis; nor is it, *pace* Marr, to produce a description. Furthermore, while it was a mistake dating back to Aristotle to suppose that there must be a *sensus communis* to unify the data received from the several senses, it is equally mistaken to suppose today, as many neuroscientists do, that the brain must solve the binding problem of unifying the information from the various sense organs into an image. If there is a binding problem, it is neither a problem of producing a unified image in the brain for the brain to apprehend or see, nor one of producing a unified image for a human being to see.

Just as the Cartesian conception of the relation of mind and body generated an insoluble problem of explaining the character of voluntary movement, so too brain/body dualism recapitulates its ancestral confusions. For voluntary movement is conceived (e.g. by Libet or Frith) to be movement caused by the antecedent decisions of the brain (precisely parallel to Eccles' conception that it is movement caused by the mind's operating on the pyramidal cells of the motor cortex). But it is an error to suppose that it makes any sense to speak of the brain's deciding anything. It is equally erroneous to think that voluntary movement is a movement caused by an antecedent act of will, a

desire, an intention or a decision.

Classical dualism advanced a conception of human nature that was constructed on a false dichotomy of inner and outer. Brain/body dualism rejected the conception of the inner as the domain of the mind, and attempted to reconstruct it as cerebral. It retained the misconceived classical conception of the body as a machine (rather than as a living organism with a good, that can flourish or suffer illnesses and senescence), differing from classical dualism in holding that machine to be driven by the brain, rather than by an immaterial substance – the mind. This cast the problem of other minds in a new light. For now it became problematic to resolve the question of why other brains should be conscious at all. That other people have brains in their skulls seems relatively unproblematic – but that their brains should be conscious seems an unverifiable and altogether redundant hypothesis. For it seems that the brain could function exactly as it does, and produce precisely the same movements it generates, without any consciousness whatsoever. Hence the question 'What is consciousness for?' seems to make sense, and speculation on the evolutionary value of consciousness ensues (cf. Barlow, Humphrey). What is awry is the conception of consciousness that informs this speculation – consciousness conceived as an inner light that illuminates what passes in the mind, like a scanning mechanism (which may have a neural correlate (as Weiskrantz supposed), or as the ineffable subjective qualitative character of experience that distinguishes human beings from 'zombies'. But if one asks what is the evolutionary advantage of being periodically awake as opposed to asleep, in a torpor, or unconscious – the foolishness of the question is patent. And if one asks what is the evolutionary advantage of being perceptually conscious of things, i.e. susceptible to having one's attention caught and held by objects or events on the periphery of one's perceptual field, the answer is obvious.[10]

[10] For a detailed discussion of neuroscientific 'brain-body dualism', see M. R. Bennett and P. M. S. Hacker, *Philosophical Foundations of Neuroscience* (Blackwell, Oxford, 2003).

6. *Behaviourism*

Classical *ontological* behaviourism was advanced by Watson, and later by Skinner, in response to introspectionist psychology. Psychology, it was claimed, is not the science of consciousness, but rather a science of human behaviour, human behaviour being understood as bodily movements. Not only is consciousness not the subject matter of psychological science, but the very concept of consciousness is useless. The committed behaviourist, Watson declared, will drop from his scientific vocabulary all subjective terms such as 'sensation', 'perception', 'image', 'desire', 'purpose' and even 'thinking' and 'emotion'. For these are names of fictions. For psychology to mature into a science, it must confine itself to what is observable, namely: behaviour. Its explanations must rest on functional dependencies between observable data, viz. environmental stimuli and bodily movements (including changes in respiration, blood pressure, etc.). Behaviourist psychology aims to discover laws correlating external stimuli and behavioural response. Speech is to be explained not by reference to thought, but by reference to causal conditioning.

Logical behaviourism was a more sophisticated doctrine. Its proponents were primarily philosophers (e.g. Carnap, Hempel). They did not argue that the mental is a fiction. Rather, they argued that the mental is reducible to the behavioural. So propositions ostensibly about mental states, events or processes are translatable without loss of meaning into propositions about behaviour and dispositions to behave.

Wittgenstein, it seems, was acquainted with behaviourism primarily through having read what Russell wrote on the subject in his *Analysis of Mind.* In some respects, he was evidently in sympathy with it. In others, he was dismissive.

Logical behaviourism was right to see an internal relation between mental attributes and behaviour, *properly construed.* For the criteria for ascribing mental predicates to other people consist of what they do and say. Behaviourists were right to emphasize that language learning is based on training, and that it presupposes shared behavioural reactions and responses. They were right to see language acquisition as learning new forms of behaviour – learning how to do things with words.

Wittgenstein agreed with behaviourists in handling understanding largely in terms of the behaviour that warrants its ascription, rejecting the conception of understanding as an inner state from which performance flows. Behaviourists were right to see uses of language as modes of action. Finally, Wittgenstein agreed with the behaviourists that psychology is not the study of an 'inner realm' that lies 'behind' behaviour, which is problematically inferred from 'external' behaviour.

Nevertheless, logical behaviourism was mistaken to think that mental attributes are reducible to behaviour. One can simulate and pretend, and one can think or feel without showing what one thinks or feels. So it makes sense, in certain circumstances, to describe someone as manifesting such-and-such behaviour but to deny that the person is in the corresponding mental state, and it makes sense to ascribe thoughts and feelings to a person even though they are not exhibited in behaviour. Similarly, it was correct to view avowals of experience as a form of behaviour, but *what* is expressed when someone gasps 'I have a pain' is not behaviour, but pain – which is a sensation. And sensations are not forms of behaviour.

Ontological behaviourism was a cruder affair. Like formalism in philosophy of mathematics, it denied the existence of something (numbers, in the case of mathematics, the mental in the case of ontological behaviourism) in order to escape from a confusion (Platonism and dualism). But the way to escape from conceptual confusions is to expose their roots, not to make ontological counter-claims. What the philosopher must do to clear up the confusions is elucidate the peculiar grammar of psychological predicates – which Wittgenstein did.

His criticisms, however, went much further. First, behaviourists misconstrue human behaviour, presenting it as mere bodily movement, rather than as *expression*, and as *purposive*, as well as *rational*, *action*. Secondly, behaviourism altogether fails to account for the intentionality of a range of mental attributes. For it takes the relationship between desire and its satisfaction, on the one hand, and belief and what makes it true, on the other, to be external. But if A wants M, then it is a conceptual truth that getting M is what is called 'satisfying A's want'. If one argues, like the behaviourist, that what A wants is what terminates a state of unease that constitutes a desire, then one

is committed to the absurdity that if a punch in the stomach terminates one's desire for an apple, then what one wanted was a punch in the stomach. Finally, although behaviourists rightly construed the use of language as a form of behaviour, they were mistaken to think that one could give an adequate account of language and of reasoning in terms of stimulus and conditioned response. The grammar of a language is a normative structure and speaking a language a normative practice.

7. *Cognitive psychology: representations and computations*

The heyday of behaviourism was in the inter-war years of the twentieth century. It was succeeded by the so-called cognitive revolution in psychology in the 1950s and '60s, which reacted to the excesses of behaviourism. Its initial aim was to revive the study of cognition and cogitation in experimental psychology. However, it rapidly abandoned this objective.[11] For the explanatory models that dominated the discipline were based on information-processing, hypothesized internal representations, and computational operations on internal representations. Here, I think, Wittgensteinian reflections can pinpoint confusions.

It is common among psychologists and cognitive neuroscientists to speak of internal representations in the brain. In so far as 'representation' signifies no more than a causal correlate in the brain of an external stimulus, this is innocuous. But it is evident that all too frequently it is meant to signify a *symbolic* representation.[12] And it makes no sense to speak of semantic (symbolic) representations in the brain.

---

[11] As is attested by one of the founding fathers of cognitive psychology, Jerome Bruner, in his autobiography *Acts of Meaning* (Harvard University Press, Cambridge, MA, 1990), p. 137.

[12] To give a few select examples: the psychologist J. P. Frisby writes 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' (*Seeing: Illusion, Brain and Mind* (Oxford University Press, Oxford, 1980), p. 8); the neurophysiologist C. Blakemore writes that maps in the brain 'play an essential part in the representation and interpretation of the world by the brain, just as the maps of an atlas do for the reader of them' ('Understanding images in the brain', in H. Barlow, C. Blakemore and M. Weston-Smith eds., *Images and Understanding* (Cambridge University Press, Cambridge, 1990), pp. 265f.; and D. Marr, an engineer, writes: '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' (*Vision, a Computational Investigation into Human Representation and Processing of Visual Information* (Freeman, San Francisco, 1980), p. 3).

Alongside the tendency to ascribe semantic representations to the brain was a further questionable commitment, namely to the idea that the brain *operates* upon representations, transforming them according to rules. This idea seems to be inspired, at least in part, by a misapprehension and misdescription of information processing by computers. But it was greatly strengthened by Chomsky's linguistic theories in the 1950s and '60s. According to Chomsky, to understand an utterance 'the mind/brain must determine its phonetic form and its words and then use the principles of universal grammar and the values of the parameters to project a structured representation of this expression and determine how its parts are associated.' To understand a sentence, Chomsky averred, is to interpret it 'by a computational process of unconscious inference'. 'The computations involved may be fairly intricate ... But since they rely on principles of universal grammar that are part of the fixed structure of the mind/brain, it is fair to suppose that they take place virtually simultaneously and beyond the level of possible introspection.'[13]

A great deal needs to be said about these misconceptions, but I shall restrict my remarks to four brief but very general points that, if correct, undermine these various computational theories.

First, it makes no sense to speak of symbolic or semantic representations in the brain. For such representations are determined by conventions. They are representations only in so far as they have a rule-governed use, and hence only in so far as there is a correct and incorrect way of using them. For an object, a sign, to be a semantic representation of anything, it must have a meaning. It is not a sign *of*, but a sign *for*, what it represents. And that it is a sign *for* what it represents is exhibited in explanations of its meaning given in a symbol-employing community, in corrections of mistakes by users, and in explanations by users of what they mean by it. But brains are not members of a community, and it makes no sense to suppose that brains can be said to employ symbols. Moreover, those who use a symbol mean something by it when they use it, but it makes no sense to ascribe meaning something to the brain or its parts.

---

[13] N. Chomsky, *Language and the Problems of Knowledge* (MIT Press, Cambridge, Mass., 1988), pp. 55, 90, 136. For detailed discussion, see P. M. S. Hacker, 'Chomsky's Problems' in *Language and Communication* 10 (1990), pp. 127-48.

Secondly, the supposition that a system of rules might be 'part of the fixed structure of the mind/brain ... beyond the level of possible introspection' is nonsensical. It makes no sense to speak of an *unformulated* rule being part of the fixed structure of the mind or of the brain. Human beings may engage in rule-governed activity without formulating the rule in so many words – they would teach the activity by example and exemplification. But brains and minds do not do so. And it is unintelligible to suppose that a rule-formulation is 'part of the fixed structure of the mind/brain', unless there is writing or speech to be found in this strange organ.

Thirdly, it makes no sense to speak of the brain's following rules, just as it makes no sense to speak of a computer following rules (as opposed to producing results that accord with rules). To follow a rule is the exercise of a two-way ability to act or not to act. But neither brains nor computers have two-way abilities. But to be caused to behave in a manner that coincides with what a rule-follower would do, to be caused to generate the same output as would result from following a rule, is not to follow a rule at all. Indeed, it is to make any rule altogether redundant for the operations of the entity (brain or computer) – since mechanical necessitation has replaced normative behaviour. A medieval monk who struck a bell every hour as determined by an hour glass, was following a rule; a church clock is not. An abacus or slide rule does not follow any rules. Neither does a computer. Nor does a brain. And when they malfunction, they do not transgress rules.

Fourthly, it makes no sense to suppose that the brain engages in computations (any more, strictly speaking, than a computer engages in computations and calculations). For to engage in calculations and computations is precisely to *follow* a set of rules, which presupposes not only a two-way ability, but also an understanding of the symbolism and of the computational rules associated with it. But the brain is not a possible subject of understanding (any more than is a computer). It cannot be said to understand any symbols or to know what they mean, let alone to use symbols and mean something by their use, for brains can neither mean nor fail to mean anything.

If these four points are correct, as I believe them to be, then computational theories in psychology, cognitive neuroscience and theoretical linguistics need extensive revision.

Wittgenstein is sometimes criticized for being a philosophical quietist. Nothing could be further from the truth. For he gave philosophy *a license* to criticize scientists. He showed *why* philosophy has a *right* to interfere with empirical sciences – for its role is as a *conceptual critic*. Philosophy is a tribunal of sense, before which erring scientists can be arraigned for transgressing the bounds of sense. They can be arraigned, not by criticizing them for deviating from ordinary usage – which deviations may be wholly innocuous – but for invoking ordinary usage and then misusing the terms invoked, through misunderstanding and conceptual confusion. Philosophy is no policeman, but an impartial judge. Scientists must be condemned out of their own mouth – by demonstrating the incoherence of *their* assertions. It is not the task of philosophy to sing the Hallelujah chorus to science or to police its pronouncement. It is rather to identify conceptual confusions that are rife in science, and to eradicate the scientific myth-making, no less than the anti-scientific myth-making, that is endemic in the culture in which we all live today.*

St John's College, Oxford

* I am grateful to Hanoch Ben-Yami and to Hanjo Glock for their comments on an earlier draft of this paper.