ABSTRACT

THE MIND AND THE BRAIN

by

Daniel C. Dennett

Recent attempts to describe the relation between the mental and the physical have met with difficulties because the supposed mental events and entities met with in introspection seem to have capacities and qualities entirely unlike those of physical events and entities. The way out of these difficulties is to follow the method of particle physics; postulate the existence of entities as the result of analysis, not as its prologue. This can be done by adopting the stance that the constituent object-words of introspective reports such as "idea", "thought", and "image" need not refer to any objects whatever for the daily business of interpersonal communication about introspected events to succeed. Introspective reports are treated with a particular blend of caution and sympathy, and pending an examination of cerebral functions and entities, no position is taken on the existence of any of the objects apparently referred to by the suspect object-words.

In the course of describing cerebral function, the concept of the content of neural signals and structures is established. With the aid of this concept, cerebral mechanisms are described that could produce introspective reports in all their differentiations, without any events or entities in these cerebral mechanisms being enough like what we suppose thoughts or images to be to warrant any identification of mental and physical events or objects. Roughly, an explanation is given of what makes us suppose or believe, when we are theorizing, that there are irreducible, non-physical mental events, objects, or qualities. No room is left for non-physical events or entities, since the physical account of the production of the introspective reports is held to be complete.

The analysis of the various aspects of the introspective world involves the examination of ordinary senses of "mental" words, such as "aware", "sensation", "reason", and "intend". Non-ordinary substitutes are occasionally proposed as theoretical words where the ordinary word is found to be vague or inconsistent in application, or an amalgam of separable concepts. Thus problems formulated in ordinary language are reformulated in theoretical terms, and this reformulation amounts to a partial solution of these problems in many cases.

THE MIND AND THE BRAIN

Introspective description in the light of neurological findings; intentionality

Thesis submitted for the Degree of Doctor of Philosophy in the University of Oxford

by

DANIEL C. DENNETT

May, 1965.

CONTENTS

PREFACE

The concept of intentionality is mentioned in the title, and in various guises it is in fact a central theme of the thesis, but the word "intentionality" appears only a few times. Originally I had intended to be much more direct in my discussion of the concept, but it eventually became clear that there was not enough room for both my analysis of the phenomenon, the "aboutness", as Feigl calls it, of mental activity, and an examination of the concept as it figures in the writings of past and current philosophers. Since the word itself is controversial and has been subjected to a variety of related interpretations since Brentano reintroduced it, the use of the word throughout the thesis would have required a lengthy accompaniment of initial exegesis, critical examination, comparison of views, and with it a large collection of cumbersome footnotes.

The literature on intentionality is noted for its exotic and obscure terminology, and my intent has been to make my points, which are relevant in a general and fundamental way to the wide variety of clashing views on the subject, without becoming entangled in factions or nomenclature. The discussion of the <u>content</u> of neural states and events is in effect a physicalistic theory of intentionality, and the arguments against mental objects, which appear throughout the thesis, are directed as much against noemata and other intentionalistic objects as against sense-data and their kin.

I am indebted to B.A. Farrell and Nicholas MacIntosh, who pointed out a number of psychological and neurological naïvetés in drafts of the first nine chapters, but I do not claim their stamp of approval on all the neurological speculations. Also to Professor Harris, who read an early summary of neurological theories and gave me advice on further reading. And I am particularly indebted to Professor Ryle for his advice and encouragement at every step.

CHAPTER 1.

Introduction.

1. Taking off the Blindfold.

There is a widespread and unfortunate tradition in philosophy that the man in the street has all the empirical knowledge required for philosophizing. In most areas of philosophy this tradition has steadily lost ground to the encroachments of science, but epistemologists and, strangely enough, "philosophers of psychology", have resisted the trend. The philosophy of mind remains a bulwark of scientific ignorance, and even the sternest critics of "mentalistic" philosophy have usually supported their attacks with only everyday knowledge or findings from the laic fringes of neurology and psychology. Great advances have been made in neurology in the last decade, and the philosophy of mind must be brought up to date or it will soon join phrenology and astrology.

This philosophical tradition of ignoring science has, in the case of epistemology, an accompanying rationale, and although it is seldom stated nowadays, it is probably the defence of ignorance that would be offered by most current writers if they were pressed. H.H. Price, in *Perception*,[1]

---

1 H.H. Price, *Perception*, 1932, p. 2.

provides a succinct expression of this rationale. Our grounds, he says, for believing the physiological accounts of perception "are derived from observation, and mainly if not entirely from visual observation". But the reliability of observation is just what is at issue for the epistemologist, and "since the premises of Physiology are among the propositions into whose validity we are inquiring, it is hardly likely that its conclusions will assist us."

Price's conclusion is a non sequitur. If the epistemologist is out to justify our knowledge-by-observation, and uses the conclusions of physiology as theorems or axioms from which he deduces conclusions about their own reliability, then his argument is of course circular. But if, on the other hand, he treats physiological findings as still-to-be-justified (but highly likely) propositions, as only hints or guidelines for the development of a satisfactory justification of observational knowledge, there is no circularity, and he can be expected to be greatly "assisted" by them. And if the epistemologist is not concerned with the justification of observational knowledge (and many are not), if in other words he accepts as obvious that our knowledge of the external world is valid, then his use - even deductively - of physiological findings does not beg any question and will of course greatly advance his course of explaining - not justifying - the genesis of human knowledge. The epistemologist who eschews physiological guidelines is unwise whatever his goal, for surely any ultimately acceptable justification or explanation of knowledge will find the results of physiological research among the paragons of reliability, and if the details of the theory produced do not harmonize with physiology, there will be an embarrassing inconsistency between the justification or explanation and what is thereby justified or explained. And for the "philosophical psychologist" to ignore physiology - as many do - is sheer folly.

I see no *a priori* reason why epistemology could not be successfully carried out in a factual vacuum, but the difficulties on which epistemologies seem inevitably to founder are difficulties over the way to describe pains, illusions, hallucinations, visual fields and so forth, and as soon as the philosopher, of whatever school, sets out to give accounts of these peripheral (not strictly epistemological) matters, he runs the risk of being refuted by scientific findings. If the philosopher finds the existence of the phenomenon of phantom limb important, why should he not wish to consider the best accounts of the phenomenon? If it is admitted that the fact of colour blindness is relevant to epistemology, why should scientific accounts of colour blindness be a priori irrelevant? The contemporary epistemologist uses a vast repertoire of anecdotes about perception, normal and abnormal, without ever considering the possibility that if these are useful guidelines, scientific explanations of these might be better guidelines.

2. Physicalism.

My aim in this thesis is to solve and dissolve these peripheral epistemological problems with the aid of neurological hypotheses about the "workings of the mind". I do not presuppose that there are no such things as minds or mental events, but argue at each step that the situations or events customarily considered to involve psychic or mental events (or in some other way not straightforwardly physical events) are in fact entirely physical in just the same way as digestion or walking is physical. The onus is the "reduction" of all mentalistic descriptions to intelligible, self-sufficient physical descriptions. And the by-product of this programme of strict physicalism will be the dissolution of several persistent epistemological pseudo-problems.

It is a measure of the implicit disharmony between science and philosophy in this area that most philosophers regard all possible forms of strict physicalism as irredeemably wrongheaded, and most scientists regard all alternatives to strict physicalism as utterly implausible, if not nonsense. To the scientist, especially the physicist, it is abundantly clear that some physicalistic explanation of mind must work or the entire edifice of science will fall, an extremely unlikely prospect. The scientist's doubts about my task will concern not its ultimate possibility but its present feasibility. But as I shall argue later, if the time is not now quite ripe for a positive physicalistic theory, it will not become riper in the foreseeable future, the brain being what it is. It is the disbelief from the philosophical side that is more troublesome, and so my thesis is primarily philosophical and secondarily "meta-neurological".

My starting point is the public, physical world of natural science and everyday life, of physical bodies and brains. I do not start with or arrive at sub-atomic physics, since the explanations presented do not happen to involve anything smaller than the biological building blocks. And I do not start from the world of introspection, the ego, or my immediate experience, although that has been a traditional starting point for epistemologies. My reasons for this are not new. I am simply subscribing to the fairly recent arguments about the priority of ordinary language and the reliance of the notion of reality on allowing the public world to be the real world from the start. As Quine, for one, says: "The trouble is that immediate experience simply will not, of itself, cohere as an autonomous domain. References to physical things are largely what hold it together."[1] And, "The familiar material objects may not be all that is real, but they are admirable examples."[2] This public point of view is now widely accepted, for one reason or another, and I wish merely to announce my allegiance to it, not argue for it.

With this starting point it is clear that I cannot be embarking on the Cartesian course of justifying observational knowledge of everyday objects. By the same token it is clear that I cannot offer direct arguments against solipsism or radical Berkeleian views, a small loss. I will be satisfied to deal them a glancing blow, which is all they can ever really be dealt. It is at least a practical impossibility to refute all possible alternatives to strict physicalism, and if some opponents can be bundled off into silly but unassailable views, that is enough.

All the same, it may seem that there is a <u>petitio principii</u> in starting with the public world and arguing

1 <u>Word and Object</u>, 1960, p. 2.
2 <u>Ibid</u>., p. 3.

against mind. But if such an argument is a *petitio principii*, so are arguments from the same standpoint against ghosts and witchcraft, and also, no doubt, arguments from the alternative standpoint of immediate experience *for* the existence of sense-data or minds. If the charge is pressed and allowed, the outcome can only be that where problems of mind are concerned there is no neutral ground on which to stand, in which case the *petitio* argument lapses and standpoint must be decided from other considerations, e.g., whether everyday physical-object talk or "protocol language" is more basic, more sure, more meaningful - a matter not to be proved one way or the other. Those who choose the Cartesian, or protocol language, or *esse est percipi* path are bade farewell with no argument from me, although considerations which I will adduce later on will tend to enhance the unattractiveness of that path - without refuting its claim for adoption, of course.

Setting aside the methodological doubters and their kin, there remains the larger and more interesting class of philosophers who accept the scientific stance and yet for various reasons wish to hold that there is a separate and somehow irreducible sphere of the mental in addition to the physical. It is their ideas, in all their guises, that I wish to challenge.

One way of challenging them is to adopt the position of rigid behaviourism, but I shall not choose this course for several reasons. First, once the position of strict behaviourism has been promulgated and secured, the argument is all but over; interesting objections are ruled out of court on behaviourist grounds of "inadmissable evidence" alone.[1] So the lines of battle are necessarily drawn up at the point where behaviourism is proposed, and the standard sterile haggle ensues, bringing out little of the best from both sides.

Second, behaviouristic accounts, if over-strict or over-reductionistic, do not follow the established rules of science as is often claimed. On the one hand strict reduction of phenomena to overt behaviour simply disallows the relevance of covert, but perfectly physical, events such as neural events. One can imagine an analogous "behaviouristic" physics that reduced brittleness to if-then statements about breaking, and disallowed information on changes in crystal structure. A common failing of behaviourists is considering behaviour to be only skin deep. On the other hand, such typical examples of behaviouristic language as "x grabbed the lever" and "x refused to do y" are clearly

1 See, e.g., Quine, op. cit., pp. 71, 235.

interpretative or teleological in a way strict "motion statements" of pure science are not.[1] The rules of behaviourism are not clearly the rules of either scientific or everyday discourse.

Third, adopting behaviourism is putting on blinkers for no conclusive reason. Introspective discourse is common and undeniably useful in practice. There are many problems over its rôle and its proper interpretation, but to throw it out as in principle unreliable and unacceptable as data is to cut oneself off from an immense if dangerous fund of data.[2] There is no denying that ordinary introspective discourse is very revealing and significant - at least in practical concerns - or that theoretical introspective discourse achieves some sort of results. Intuitively, Husserl's Phenomenological descriptions, for all their turgid vocabulary, "ring bells" with great frequency and consistency, as do the introspective accounts of other phenomenalists and the best novelists. We recognize the "insights" in these accounts, and this fact alone is extraordinary enough to warrant investigating introspection, and

1 See Charles Taylor, The Explanation of Behaviour, 1964.
2 Cf. Mace, "Introspection and Analysis", in Black (ed.), Philosophical Analysis, 1950, and Burt, "The Concept of Consciousness", British Journal of Psychology, 1962.

suggests that there is something to be salvaged from the introspective idiom.

For these reasons I accept the burden of analyzing introspective accounts and rebutting objections based on them. But that does not mean that I must accept introspective reports on their face value, as reliable, referential statements about occurring inner events. My method of dealing with introspective reports will be explained in Chapters 2 and 3.

The task set out may seem wildly ambitious so it will be well for me to say what I am not doing. I am presenting a fairly detailed neurological theory of mind, making use of concepts from cybernetics, but I am not attempting to consider or rebut all the recent philosophical arguments about whether or not a mind or brain can be a machine and vice versa. Alan Ross Anderson, in the introduction to his collection of articles, Minds and Machines (1964), mentions that since 1950 over a thousand papers have been published on the question as to whether "machines" can "think". This literature is highly redundant, usually either oversimplified or bizarrely abstruse, and in general, I feel, quite futile. My position on the subject becomes quite clear in passing: anything a mind (or brain) can do, a machine could do - for what that is worth. The more direct, more interesting question, Can a brain be a mind?, is the central theme of the thesis.

I am also not attempting a "complete" epistemology or philosophical psychology, although the chapter titles might suggest this. I deal with each facet of the "mind" - privacy, certainty, awareness, perception, knowing, reasoning, intention - only to the extent that neurological and epistemological considerations and interests overlap. Thus I do not say much about emotion or the Freudian Unconscious because these are areas of scant epistemological concern (but great neurological concern). In Chapter 7, on certainty, I do not deal with tautology or logical certainty, but only with the supposed certainty of introspective reports and its physical explanation. In Chapter 12, on knowing, I isolate the psychological aspect of knowing from the metaphysical or metascientific aspect (the relation of knowledge to Truth). Reasoning, in Chapter 10, is considered only in so far as it is an occurring "mental" phenomenon in need of and amenable to physical description. This principle of overlap is relaxed only where wider argument is required to sustain the overriding theory of physicalism. The <u>pure</u> epistemological questions - concerning confirmation, deduction, certainty, etc. - are thus distilled out of the messy stew of philosophical psychology and left unconsidered, leaving room for a detailed analysis of the independent problems of mind and mental events.

It might also be well to add that my hypotheses about neural mechanisms are based primarily on philosophic and cybernetic considerations; I am not a trained neurologist. But every effort has been made to back up my hypotheses with the most recent discoveries and ideas of neurologists. I do claim that no neurological evidence has come to light that might tend to disconfirm my proposals, and in fact most of my proposals are simply extensions, with an eye to philosophy, of the most promising suggestions of recent neurology.

CHAPTER 2.

Theories and Words.

3. The Classic Alternatives.

If one starts with the traditional but suspect assumption that we all know what we are talking about when we talk about mental events, there seems to be a strictly limited number of positions one can hold regarding the relation of these mental events to physical events in the brain. That is, if one accepts the traditional formulation of the mind-body problem, plain trial-and-error shows that between these two varieties of phenomena there could be no relation, identity, or something in between. As regards intermediate views, one could hold that there was interaction between the two spheres, or not. And if not, the only kind of relation left would be some sort of isomorphism, some (partial) parallelism between the mental and physical events.

So at first blush there seem to be only four possible views regarding the relation between mind and body, granting that the skeleton of each view might be fleshed out in different ways: (1) there is no interesting relation whatever between the events of the mind and those of the brain (the only relation being, say, joint ownership by a person); (2) mental events just <u>are</u> physical events in the brain;

(3) mental events interact with physical events in the brain; (4) mental events are somehow and to some degree parallel to events in the brain.

(1) is out of the question. The uncontroversial fact that alcohol affects the brain, and that changes occur at the same time "in the mind" with law-like regularity should eliminate this view from consideration.

(3) should be similarly unattractive to anyone with any allegiance to the principles of science. The only evidence for interaction of physical with non-physical events would be as follows. A neurologist traces a certain physiological process A at time t to point P in the brain, and then traces a particular further process or event B from that point. Then he traces a qualitatively identical process A' at time t' to point P that is followed by the further process or event C, which is qualitatively different from B. In other words, he would have to discover that two qualitatively identical processes A and A' had different continuations from point P. If this were the case and only if this were the case would the neurologist have any reason for supposing that there was some non-physical event - or difference between A and A' - that caused the physical difference between B and C. But to say that A and A' are identical in physical qualities is to claim that one has total, perfect

knowledge of the physical qualities of A and A'. Otherwise A and A' might differ in some undiscovered physical way that could explain the difference between B and C. So interaction could only be posited on the assumption of total physical knowledge. Such a claim would always be outrageous, even in such restricted and simplified physical environments as billiard tables; the protean environment of the brain would rule out even the rough plausibility in practice of such a claim.

Interactionism puts the concept of the non-physical under considerable strain. The original appeal on behalf of non-physical entities is that thoughts, dreams and ideas are just so unlike any physical entities that they must be non-physical entities. Behind this dubious conclusion is the notion of the category mistake. Mental events, so runs the formula, are simply and obviously not the same <u>kind of thing</u> as heartbeats and horseraces and other physical events, and mental objects like images and thoughts are obviously very different from physical objects like paintings and tables. But surely thoughts and ideas and images are also not at all the kind of thing that could interact with an electrical discharge (A or A') in the dendrite of a neurone. Unless thoughts and ideas have some qualities that are invisible even to the introspector, thoughts and ideas are just not remotely like things that could interact with anything except perhaps, mysteriously, another thought or idea.

But if such an invisible substratum of psychic electricity is invoked as a deus ex machina (a machina ex deo?) - as is perhaps the case is some versions of the "double aspect" theory - the view lapses into anti-science or magic. For if the things that interact with physical brain processes are truly non-physical, they must be in principle undetectable by physical means. There could not be a piece of apparatus, for example, that detected non-physical differences. There might be a piece of apparatus that could detect hitherto undreamed-of differences of awesome strangeness, but these would be construed as new types of physical differences. Far better to enlarge the scope of the physical than to clutter the scientific domain with two disparate but bridgeable universes.

If the non-physical is in principle undetectable, interactionism is not a theory of mind, but a claim that there can be no theory. To posit interaction is to say: since there is a difference between events B and C, there must be a difference between A and A', and since I can discover no difference, the difference must be in principle undiscoverable. Interactionism gives up the search when in fact the search has hardly begun. The point is not that there could not be interaction, but that even if there were, unknown to us, interaction, interactionism would be a defeatist and anti-scientific view.[1] My aim is to show that such defeatism is not even tempting.

That leaves (2), the identity theory, and (4), parallelism, and these are simply two faces of the same coin. Any view that is strictly parallelistic differs from the identity theory only in a certain squeamishness about asserting the strict identity, not merely the correlation, of mental events and brain processes. Parallelism depends just as much as the identity theory does on the successful tying together of reported mental events and simultaneously occurring cerebral events. A theory that eschews this responsibility is not parallelism, but "no-relationism" or interactionism. To accept that there is some mental event but deny that it must have a physical event "corresponding" to it is to admit that there is some physical event or difference, such as a person's utterance of a true report of such a mental event, that is not explicable solely in terms of other physical differences or events.[2] So parallelism is just as dependent

---

1 Cf. Feigl, "The 'Mental' and the 'Physical'", in Minnesota Studies in the Philosophy of Science, Vol. II, 1958, p. 428.

2 Cf. Brandt, "Doubts about the Identity Theory", in Hook (ed.), Dimensions of Mind, 1960, pp. 57-67.

for its details on neurological spadework as is the identity theory.

Since the two theories differ in ontology and wording but not in confirmability or predictive power, is there any way to choose between them? The identity theorist would argue that the parallelist can say nothing the identity theorist cannot, and that the parallelist simply clutters up the conceptual scheme with unnecessary strange objects, the mental events that run parallel to the physical ones. Stressing at once the queerness of the "correlation",[1] and the uselessness to theory of the extra entities, an appeal is made to Occam's Razor, the principle that we should not multiply entities beyond necessity. As Quine says, "If there is a case for mental events and mental states, it must be just that the positing of them, like the positing of molecules, has some indirect efficacy in the development of theory ... The bodily states exist anyway; why add the others?"[2]

The parallelist's reply has usually been that the identification of mental events with brain processes is a category mistake. The argument runs: I knew I had thoughts,

---

1 Cf., Feigl, op. cit., on "nomological danglers", and Smart, Philosophy and Scientific Realism, 1963, pp. 68, 90 and 94.

2 Word and Object, 1960, p. 264.

feelings, dreams, and other mental events, and could identify them, before I knew anything about brain processes; then, since recognizing that I am having a thought does not involve recognizing that there is a brain process occurring, thoughts and brain processes cannot be the same things. But this argument is surely confused, as J.J.C. Smart has pointed out repeatedly.[1] As he says, we could all talk about lightning before we knew it was an electrical discharge, but that does not mean that lightning is something else, only parallel to the electrical discharge, nor that it is a category mistake to say that lightning is identical with certain electrical discharges. It is an empirical discovery that lightning is electricity and that thoughts are brain processes. Smart's example is perhaps unfortunate because of the quibbles that can arise over lightning being only the visible effects of the discharge, and not in fact identical with the discharge. But Smart's point can be illustrated with stronger examples. Smith may be the Mad Strangler, unbeknownst to me, and I may see Smith. If

---

1 See Smart, op. cit., and "Sensations and Brain Processes", Phil. Review, 1959, pp. 141-56; and the replies by Stevenson, Phil. Review, 1960, pp. 505-10; Pitcher and Joske, Australasian Journal of Philosophy, 1960, pp. 150-60; and by Baier, ibid., 1962, pp. 57-68. Smart replies in Phil. Review, 1961, pp. 406-7; Austral. Journal of Philosophy, 1960, pp. 253-4, and 1962, pp. 68-70.

I say "I have seen Smith", I am not saying I have seen the Mad Strangler. If I then find out (an empirical discovery) that Smith is the Mad Strangler, and someone then asks me if I have seen the Mad Strangler, the answer is Yes. I may have been talking quite successfully and correctly about Smith for years, without knowing that he was the Mad Strangler, but if someone then shows me that Smith is the Mad Strangler, it would be absurd to say: "But you are making a category mistake. I have known Smith for years, and that did not at all involve my knowing he was the Mad Strangler. Smith is merely parallel to the Mad Strangler."

If this argument of the parallelists is obviously misconceived, the misgivings that produced it do not vanish once the argument is refuted. An alternative expression of these misgivings is that mental entities have qualities that brain processes do not. The most felicitous expression of this argument I can muster is: when I perceive a tree my mental image has extension, colour, shape (big, green, tree-shaped), and these qualities are quite different from the extension, colour, and shape of a brain process (small, gray, network-shaped?). The difficulty with this argument is that there are many commonplace and important questions that cannot then easily be answered: How large is a mental image? Is it three-dimensional? Where is it? No one has yet been able to present a trouble-free account of the spatial or quasi-spatial characteristics of mental images, and as I shall show later, the task is fundamentally misconceived. A more direct way of adjudicating between the views is to drop the initial supposition that we all know what mental events and objects are. A course then lies open for a compromise that avoids parallelism's inflated ontology and weasel word, "correlation", and the identity theory's antagonizing claims of strict numerical identity.

4. Physicalism without Identity.

From this new point of view, while it is certainly wrong to say that someone's reports of thoughts are meant as reports about brain processes - something about which he may know nothing - it may also be too strong to say that if someone reports that he has just had a thought, the word "thought" refers (unbeknownst to him) to the same entity referred to by some description of a brain process. Feigl would have it that "the 'mental' states or events ... are the referents (denotata) of both the phenomenal terms of the language of introspection, as well as of certain terms of the neurophysiological language."[1] But from the fact that

1 Op. cit., p. 447. See also Smart, Philosophy and Scientific Realism, pp. 92-105.

introspective reports of mental activity are undeniably significant as whole utterances, it does not follow that the terms in these utterances which are _prima facie_ (grammatically) in referential position are genuinely referential. The move from the general significance of an utterance to the desirability of considering its constituent terms as referential is a natural move, but not necessarily a wise move.

Suppose the idiom in ordinary language for reporting that one is tired were "I have an exhaustion", and instead of saying "I am more tired than you" one said "I have a more intense exhaustion than you". The analysts of ordinary language could be relied upon to have examined the concept of exhaustion, and they would have discovered that I cannot have your exhaustion, that the state of having an exhaustion is not so easily datable as the state of having a sixpence in one's hand, and that unlike pains, I can have only one exhaustion at a time. If this were the case, it would be at least misleading if scientists eventually discovered what physical differences there are between tired people and others, and then announced that exhaustions were in fact _identical_ with certain shortages and lowered capacities in our bodies. Rather the entire report of exhaustion should be replaced by an account that the body is different in these ways when one truly reports that one has an exhaustion.

It may be argued against this that we have better reason for our use of the "I have x" form instead ofbthe "I am x" form in the case of reports of thoughts than in the case of reports that one is tired. This is not denied, if it means that it was more <u>natural</u> for us to develop the "I have x" idiom for mental activity reports, but naturalness is not a conclusive reason for permanent fashions in reification. Of course it was more natural for us to develop an idiom for certain mental activities as <u>things had</u>, and this can be explained by the use of a little myth, without calling up psychic entities.

The idiom "I hear a bell" came into existence originally in the days of our primal fathers. They had things called bells, and when they struck these things there was something unusual (not necessarily the existence of some thing) that they could report. They developed the idiom: "I hear a bell". Later they found they could report much the same event when out of sight of the bell being struck, and they said, naturally, "I hear a bell". So they developed the idiom "to hear x" where x was some <u>object</u>, like a bell, a bird, a falling stone. Then they came to realize that there were refinements in this. Sometimes they heard only part of an object, so to speak. They could hear that a man's mouth was making noise, but not that his running feet were - since they were too far away to hear his feet." So they kept the verb/direct-object form, and added, by analogy, a new object, a voice. Then they could say: "I hear the man's voice, but not his feet." When this came about, they had a new and strange type of object. It was not a man's mouth, or his lips, or his lungs and larynx, or even all of these together; nor was it an event like a footfall - since a man had his voice whether or not he was exercising it. It was simply the object one could hear when a man did certain things with his mouth and lungs and so forth. Our ancestors, not being philosophers, probably did not notice their category confusion, and besides, the idiom worked well, seldom if ever leading to misunderstandings. Fortunately for them, no philosophers came along who were interested to ask whether voices were identical with certain bodily processes or parts, or merely parallel to them.

Later they came to realize that sometimes there was another kind of report they wished to make. Their experience was rather like hearing, but not like hearing (the collisions of) everyday objects. They kept the going grammar and posited new things called "thoughts", and said "I hear a thought". Soon it became clear that what was happening was not really just like hearing - it happened even if you blocked your ears, and to some extent it seemed you could hear what you wanted to hear - so they made their idiom more general and said "I am having a thought". Or, they said "I am thinking a thought". (It would be a misuse of etymology, of course, to suppose that the concept of thinking preceded the concept of a thought, or a similar concept, just because "thought" is a derived form of the verb. It really makes no difference which came first.) This idiom worked well. If someone wanted elucidation he had merely to ask "What is your thought?", and he was informed, whether or not there actually was such a thing as a thought. (One can elucidate with "I have a tenor voice", just as if there were such a thing as a voice.)[1]

Now just as the physiologist is not required to accept the ordinary reification of voices, so the philosopher is not required to accept reification of thoughts and ideas. The physiologist in fact should not say that a voice is identical with a complex physical situation, but that if

1 This myth is not intended to show that it was logically necessary that the introspective idioms were formed by analogy, but just that it makes good sense that they were. See Geach, Mental Acts, 1957, pp. 20-1, for an argument - not espoused by Geach - that such use of analogy in coining introspective idioms was not necessary. In fact it is hard to imagine how non-analogical coinages could ever have gained currency. How does the imaginative ancestor, who coins the highly original idiom "to me gleeb" instead of "I hear a thought" or "I have an idea", make himself understood, except by analogy?

someone is described as having a good voice or a tenor voice or no voice, then his physical condition is such and such, <u>and</u> this physical situation is all there is to having a good, tenor, or no voice. I shall argue that the philosopher should follow the example of the physiologist. He should not say that thoughts or other mental entities are <u>identical</u> with brain processes, but that if a man reports that he is having a particular thought, then there must be some particular brain process occurring, <u>and</u> there is nothing else involved when one is said to be having a particular thought.

For the moment I am proposing a course of action, not a view. I am leaving it open whether it is desirable to treat the ostensibly referential introspective words as referents for theory. Following the good example of method in nuclear physics, the initial <u>explananda</u> will be uncontroversial physical phenomena such as apparatus readings and human utterances; more unusual entities will be posited along the way and if and only if they are required for theory.

On this method, introspective reports will be treated, like the loud "reports" we call automobile backfires, as mere <u>explananda</u>, and not as necessarily significant at all; they are given the rôle of symptomatic noise - the same rôle accorded by physicians to crying, belching, and panting.[1]

---

1 Brandt, <u>op. cit.</u>, in Hook, <u>Dimensions of Mind</u>, approaches this course in correlating brain processes with "associated

The only difference acknowledged from the start is the immense complexity of this noise: it mirrors the grammar of stable, things-in-the-world speech, and is replete with hundreds of <u>conventional</u> variations. To the objection that when one says he is having a thought about China, he <u>means</u> what he says referentially, and not as symptomatic noise, the reply is: that may be, but for the time being such utterances will be taken as noise only; there is no denying that uttered statements are complex noises.[1]

It should be stressed that I am not adopting Wittgenstein's view that "the verbal expression of pain replaces crying and does not describe it" - and hence is not a referential report.[2] Nor is this Ryle's view that such reports are "avowals" and not assertions.[3] Ryle has since described to me a hierarchical view with "Ouch" at one, avowal, end; and "the pain is in the third tooth, upper left" (as said to the dentist, for example) at the other, reportorial, end.

---

mental facts", but here he is equivocating between the fact as event (such as might be referred to in a report) and the fact as the report itself. See also Gustafson, "On the Identity Theory", <u>Analysis</u>, 1963, pp. 30-2, who talks of "the <u>fact</u> that ..." and seems to equivocate between situation or event and something like proposition.

1 Cf. Putnam, "Minds and Machines", in Hook, <u>Dimensions of Mind</u>, p. 170.

2 <u>Philosophical Investigations</u>, i 244; see also i 367, i 370.

3 <u>Concept of Mind</u>, p. 102.

I am also neither accepting nor rejecting this view.[1] I am not saying that such utterances are or are not genuine reports; I am not taking them as reports, pending an examination of cerebral processes.

It should come as no surprise if the ordinary introspective idioms turn out on examination to be only metaphorical and their constituent terms poor referents. This has already been suggested indirectly by Ryle, when he asks his embarrassing questions about the putative entity-words of ordinary mental-activity reports. How many volitions are required to tie a square knot? How many thoughts are involved in thinking of one's family? Are some ideas bigger or longer than others? Where are pains? How large are mental images, and where are they? The fact that these words work so well in their ordinary language contexts is no evidence that they are reliable referents; on the contrary, their manifest dependence on limited, conventional contexts reveals their dubious value as referents.

Referentiality need not be seen as absolute in any way. A case can be made that all words are contextually dependent to some degree, and hence all language, and all conceptualizing, might be called metaphorical. If this move is made,

---

1 Smart reluctantly rejects these views. Philosophy and Scientific Realism, p. 91.

however, "metaphorical" loses its utility. If all reification is conventional in that language is in the end a social creation, reification that allows one's referents the greatest contextual scope will be the most fruitful, the least misleading.[1] Treating a term as non-referential is not giving it a zero reference; "dodo" is referential, but its reference is empty; "mile" is just not referential.

Once it is determined just what sorts of events and entities there might be in the brain, it will be seen that explanations of the supposedly non-physical mental phenomena can proceed without positing any objects or events enough like what we generally suppose mental objects and events to be like to warrant an identification. This will militate against the identity theory, but not <u>for</u> some parallelistic alternative. Rather, the completion of an adequate explanation involving no entities for the mentalistic words to refer to militates against the view that these words should be taken as referential. The explanation of vocal sound production by the physiologist does not show then since no physical object or condition is well referred-to by "voice", there is an additional parallel

---

1 This is much the same argument as Quine's, in <u>Word and Object</u>, Section 50, that ontic commitment is a matter of what to admit as values of variables under quantification, and that such undesirable referents as "behalf" and "mile" are to be smothered in their limited contexts as unanalyzed and "bundled off" into relative terms.

object or condition, a non-physical voice; it shows that "voice" is non-referential.

This course, if successful, will avoid the problem of the relationship between mental entities and physical entities by not including mental entities in its initial ontology. Whereas Smart, for example, eventually repudiates non-physical mental entities, his difficulties arise from accepting mental entities as objects for discussion before telescoping them into brain processes. He is like the physiologist who rightly argues that there is nothing non-physical about a voice, but then goes on to say just what physical thing a voice really is.

CHAPTER 3.

Privacy and the Subject Veto.

5. Undeniable Final Authority.

Any physicalism must rule out the possibility of necessarily or intrinsically private mental events, states or features. If the entire range of "mental" phenomena is to be reduced to physical events and states, then since these physical events and states are objective and are held to determine all human behaviour in all its details, "mind reading" on the basis of these physical events and states must be held to be possible in principle.

Mind reading for the physicalist would involve having the final authority on the nature in the broadest sense of a person's reports of mental experience. We ordinarily make distinctions among our own and other people's reports of mental experiences; we say that some are certain, some uncertain, some are lies, some are poorly expressed. Any convincing theory must explain or explain away these distinctions we so confidently make. We are confident that having a mental image of the Eiffel Tower and having a mental image of Magdalen Tower are different phenomena, and although this way of describing them may be misleading, the difference in virtue of which we make and apparently understand these different accounts must be described.

The possibility of such mind reading is often denied in one way or another by philosophers, so before proceeding with the details of a positive physicalism it is well that arguments about possibilities in principle be cleared up first. Ayer has recently presented a case for a limited form of necessary privacy. If even his very cautious case for necessary privacy can be rejected, bolder claims that the proposed task is impossible in principle can be headed off.

> When it comes ... to a person's knowledge of his present thoughts and feelings, then I do think that there are many cases in which we logically are obliged to give him the last word. Even if we allow it to be possible for others to become aware of his thoughts and feelings in the way he does, their knowledge of them will be subordinate to his. The accuracy of their reports will be checked by his, and where there is disagreement his verdict must prevail. Thus, even if one's mental states are not private in the sense that there is any single way in which, of necessity, they are detectable by oneself alone, they may still be private in yet another sense. One may be the final authority concerning their existence and their character.[1]

Or in other words one may be the final authority concerning the character, e.g., sincerity and accuracy, of one's reports of mental experience. Ayer does not consider just why we are logically obliged to give the subject (the maker of

---

1 Concept of a Person, 1963, p. 68.

the reports) the last word. I shall argue that at most there is an obligation to grant this final authority which is shared by members of the Ordinary Language Users Society - all of us, in other words.

Our ordinary stock of mental words is set aside from the rest of ordinary language by a peculiar difference in application. Whenever we apply one of these words to another person, we grant the person to whom the word is ascribed the final authority as to the correctness of the particular ascription. This final authority proviso I call the <u>subject veto</u>, because although we may have strong reasons for holding that, say, Smith is thinking (or not thinking) about apples, Smith's word in the matter is final; if he insists we are wrong then all we can do is believe him. We may not be wise in doing this, but the rules of ordinary language bid us to do this.

The use of the subject veto is inculcated in the process of learning mental words. As children we learn to ascribe words to things, and this ascription is anchored in intersubjectively observable differences and similarities in the world - especially in observable behaviour.[1] But with mental words, although we are <u>guided</u> in our ascription of them to

1 Cf. Quine's discussion of language learning in <u>Word and Object</u>, chapter 1.

others, and initially to ourselves, by behavioural data, we also learn as a fundamental step toward getting the words right to concede the last word to the subject of the ascription. Until the child has learned that the behavioural clues are only hints, and that one has to ask a person to be sure, he will not be using words like "want", "think", "pretend", and even "see" correctly. No doubt most of this conditioning is achieved in the child without ever becoming explicit in his training, and no doubt it is aided by a great deal of possibly erroneous theory given him by his parents: "I have a picture of something in my mind, Johnny; guess what it is", "What do you want now? I can't tell what you want". The details of this inculcation, however interesting, are not important here; it remains that the child is conditioned from the beginning to accept the subject veto. He learns, by learning not to say such things, that to insist that one knows better than the subject does what he is thinking or experiencing is simply deviant.

It is not my purpose to show that all or only mental words (whatever they might be) exhibit the subject veto, nor to show that the subject veto is ironclad. To suppose that would be to suppose much greater regularity in ordinary language than would be credible. In fact the domain of the subject veto has many important penumbral cases and many exceptions. Consider the verb "to love". If a young man protests "but I <u>do</u> love you!", but is virtually belied by his actions, no wise girl will grant him the subject veto, and yet most of us will accept, from the same young man, such statements as "<u>I'm</u> the only one who can tell if I love her". This very real confusion over the status of the subject veto of "love" could only be settled by arbitrary legislation. The word is, as literature suggests, fundamentally ambiguous. "Remember" presents a different kind of fluctuation. The subject veto is not allowed in cases like "I remember the dates of the Tudor monarchs", which can be tested, but is allowed, under certain biographical conditions, after such questions as "do you remember it (as an experience), or have people simply told you about it, or are you imagining it?"

Many rather non-mental words allow the subject veto. There is Ryle's example (quoted by Ayer in <u>Concept of a Person</u>, p. 15) of the child who says he is drawing a ship and cannot be contradicted, no matter what the scrawl looks like to the psychoanalyst or the art critic. The so-called intentional verbs, like "hunt" and "want" also display the pressures of the subject veto. And conversely, the undeniably "mental" verb "to know" very seldom allows the subject veto when a person professes knowledge, but often allows the subject veto when ignorance is professed. ("I didn't know the gun was loaded.")

The subject veto is at its best where there is little or no behavioural data to suggest the answer, as in the case when someone looks at Wittgenstein's duck-rabbit and says he sees a duck. It is from these cases that the subject veto derives its force when it is applied to more dubious cases. The child drawing the ship is a strong case, but when a painter draws a woman and calls it a rose, we are in doubt about what to do. There is a principle of charity, or perhaps patience, that determines how far we let the subject veto go; even in preposterous cases the hypothetical presence of the subject veto will be felt by the most uncharitable observer. It is not denied that people sometimes do deny the subject veto, but even when they do, there is a lingering doubt about whether such denials can ever be more than firm opinions. One feels there's always room for the subject to hold to his claim, and that this room can never be appropriated by any outsider.

Tracing in detail the applications and exceptions of the subject veto for various words is a lexicographer's interest, if anyone's. The point is that the subject veto's application, including its real or imagined inconsistencies, is learned and tacitly accepted by all speakers of ordinary language. This acceptance by the language-learner of the subject veto is not to be separated from his acceptance of a theory of privacy of mental events or acts. That is, an indispensable part of learning to use these words in agreement with society is the learning of a theory of mind embodied in the application of these words. This theory is a part of our conceptual heritage, as much engrained in our language as the theory of enduring physical objects.[1] It is clear to a certain extent what the natural, inherited theory is: people do things, have things, feel things, that are absolutely private, and, as Ryle says, "People tend to identify their minds with the 'place' where they conduct their secret thoughts."[2] One might call this "folk philosophy". Only the details that are strongly assumed by ordinary language need concern us. It would be unwise to try to construct a more sophisticated theory of mind from the "logic" of our mental words. Such a theory, ex pede herculem, would have as it's only support that generations of man had presupposed something like it in their ordinary

1 It may be objected that usage in ordinary language cannot by itself embody theory; an interpretation must be put on usage before there is any theory. But ordinary language can certainly embody directions toward "natural" interpretations, so that given certain features of ordinary usage, one naturally puts interpretations on it so that a particular theory emerges. More strongly, I would hold that there can be theories (my own being one of them) that cannot be expressed in ordinary terms without these terms being deviantly interpreted.

2 Concept of Mind, p. 27.

talk, with or without realizing it.

Whether or not this folk philosophy is right - that is, whether or not the subject veto is a useful or proper feature of language - it is with us, an empirical fact about ordinary language that has philosophical implications. For one thing, the criterion hunt initiated by Malcolm and others should be called off. If in fact the subject veto is in force, then the search for airtight behavioural criteria for the use of pain-expressions, for example, is misguided. The sentence "he is in pain" allows the subject veto. This means that behavioural data serve as hints or clues but not as final criteria. They become more like criteria the more convincing they are in individual cases, but the subject veto is always allowable in principle. Whether or not this is as it should be is another question, and if anyone concludes that there <u>should</u> be behavioural criteria for such words as pain-words, then he can legislate on the meanings of new, non-ordinary words. But any attempt to anchor the ordinary meaning of pain-words or other subject veto words to behavioural criteria is flying in the face of the obvious facts about how we learn to use these words correctly. Quine suggests that we must all be, to begin with, objectivists, in that we perforce inherit the conceptual scheme of objective physical reality in learning to speak, and this conceptual scheme must be the <u>starting point</u> of any comprehensible science.[1] But by the same token we are mentalists, not behaviourists, in that the concept of a private mental theatre is embedded in the workings of our mental-word vocabulary.[2]

6. <u>How to Deny Final Authority.</u>

In general then, those who would deny the occurrence of private mental events must step out of ordinary language. This holds for the philosopher, the psychologist and the neurologist.[3] The human brain is so immensely complicated that it is unlikely that neurologists will ever succeed in translating (in a stretched sense of that word) all the high-level neuronal signals, and so it is very unlikely that methods of neurological mind reading will ever be perfected. The success of such an attempt would probably not be worth the effort and cost. But the way in which such an endeavour would have to proceed sheds light on the question of the possibility of abandoning ordinary language.

---

1 <u>Word and Object</u>, §1.
2 That ordinary language implies a dualism is argued by Shaffer on different grounds, in "Could Mental States be Brain Processes?", <u>Journal of Philosophy</u>, 1961, pp. 813-22.
3 Consider, for example, Freud's concepts of unacknowledged desires, submerged reasons and motives, and unlovely loves. By disallowing the subject veto, Freud certainly uses these ordinary terms in non-ordinary ways.

The equipment that would be required by the neurologist for such a task would be many times more sophisticated than present apparatus. Ordinary electrodes and encephalographs would not suffice as data-gatherers since they detect only the gross activity of the brain. The successes that have been achieved in deciphering the functions of individual neuronal signals have depended on vivisection and micro-electrodes, and since the use of this method on a wide scale must be practically ruled out, some other method would have to be devised for producing a detailed cerebral map on which all the individual neural signals could be traced. Then, the speed and complexity of neural function are so great that super-computers would be needed to evaluate the cerebral map.

The purpose of such an apparatus would be to explain the occurrence of the _explananda_, behaviour in general, in terms of the occurrence of cerebral activity bound by the laws of physical science in such a way that, given the description of the neural activity, the description of any _explanandum_ is uniquely determined. Actual prediction, in the sense that the computer predicts at time t the occurrence of behaviour x at time t + a may be ruled out by the practical inability to control all the input factors (sensory stimulation) that contribute to neural activity, and the inability of the computer to operate fast enough to beat the brain with its evaluation of the outcome. It is, however, the predictive power of the theory and not the hardware that is important.

There are two ways in which the neurologists might begin their analysis. They might begin by searching for patterns in neural activity, classifying these patterns, and then developing a theory to explain the interaction of these classes. The theory would consist of statements like "A-patterns interact with B-patterns to produce X-activity in the speech centre". Such a course would avoid ordinary language descriptions of mental activity entirely. But such a theory, however consistent and predictive, would be of little application unless further correlations were made, like "whenever there is X,Y,Z-activity in the speech centre, an utterance is produced that is called, in ordinary parlance, a report of imagining Paris in the spring." Without links with the domain of ordinary discourse about behaviour and mental activity, the knowledge of the interaction of neural patterns could not be put to any non-neurological use, unless new, non-ordinary ways of talking about human behaviour were also developed to coincide with the neurophysiological findings.

The alternative starting point for the neurologists would be first to classify utterances and other behaviour according to the differentiations of ordinary language and then to search for similarities in neural patterns that regularly occur when a particular type of behaviour (as ordinarily considered) occurs. Both programmes could arrive at the same body of explanations, and each has advantages. The first would be less vulnerable to any misleading preconceptions embedded in ordinary discourse; the second would provide useful but not necessarily reliable hints about where to look for similarities of function in the brain.

In making correlations of neural situations with ordinary parlance differentiations, both programmes must accept, at first, the operation of the subject veto. In isolating the class of sincere reports of imagination, for example, the neurologists must accept the subject's characterization of a group of utterances as such, since the subject's characterization is the final criterion of the class in ordinary parlance. While no commitment would be made on the referentiality or significance of introspective reports of imagination or experience, statements by the subject <u>about</u> such utterances could be accepted as <u>bona fide</u> referential statements. A conventional lie-detector could be used in practice to screen out malicious interference by the subject, but the reliability of lie-detector results should not be an incorporated principle of the theory, since this would pre-judge at least part of the question of privacy.[1]

Now suppose that a successful theory were established, by either programme, with regular correspondences of neural patterns with types of utterances and other behaviour, including systematic correspondences of particular neural situations with reports of particular experiences. For example, there would be well-confirmed statements like "the neural situation x,y,z, ... always produces a sincere report that the subject is thinking fondly of his mother and sees her in his mind's eye smiling and with her hands folded in her lap ... ", and "the neural situation x,y,q, ... always produces the same sincere report except that the mother is crying". Suppose a programme has been written for the super-computer, so that the coherence and predictive power

1 For ease of presentation, these scientists could use the ordinary idioms as handy referents for physical states of affairs, all the while considering them as semantic atoms. Thus they could speak of the subject as having-a-mental-image-of-Paris if this is how he describes his state, without considering the idiom as anything more than a tag-name. Similarly, the radio repairman could talk (if it happened to be convenient) of there being a beep or a buzz in a speaker assembly without thereby treating the "reports" of the speaker, "beep!" and "buzz!", as referential or significant. This relaxed usage will be followed wherever it is convenient and safe in subsequent chapters.

of the theory can be conveniently tested. The computer is humming along, predicting descriptions of visual experience, reports of fatigue, occasional fibs, free associations, and so forth - <u>with never a slip</u>. If the subject recites a poem in his mind and never utters it aloud, the computer reels off the poem, complete with accounts of associated imagery, stumbling points, "tone of voice", boredom with the experiment and so forth. Over a fairly long period of time, let us suppose, no discrepancies occur; the subject is shown the computer's results regularly and uniformly vouches for their accuracy.

When such success is achieved, the neurologist may declare the subject's stamp of approval no longer necessary and proclaim the unqualified ability of his apparatus to read minds - or at least the mind of his one subject. To borrow Wittgenstein's phrase, having climbed the ladder of the subject veto, he may now throw it away. To cling to the requirement of subject approval would be to posit indirectly the existence of an unknown (and perhaps unknowable) factor operating on the <u>explananda</u> when the known factors had been found, empirically, to determine the <u>explananda</u> uniquely. This would be like holding that there are poltergeists responsible for the operation of internal combustion engines, even though their operation has been adequately described and accounted for by physics.

If the success of the predicting apparatus was not perfect, on the other hand, if there were frequent and regular discrepancies (relative to the subject's approval), the neurologists would be sent back to the drawing boards. Depending on the degree of success achieved, they would try to salvage as much of their theory as possible, and would start hunting for previously discounted or undiscovered factors. They would not adopt interactionism, although in their darkest moments they might entertain the idea. But if there were only, say, one recorded discrepancy in many hundreds of hours of experimentation, it would be irrational to reject the theory or to relegate it to the status of "unproven". Even if the discrepancy could not be explained away as an apparatus breakdown, it would not affect the hegemony of the theory, provided that the same or a similar discrepancy could not be reproduced regularly, in which case there would be "frequent and regular discrepancies". Smart, stressing that his identification of mental events with brain processes is contingent, says, in discussing a similar imaginary experiment, that <u>one</u> disconfirmatory result would cause him to abandon the brain process theory.[1] But

---

1 Op. cit., p. 99. See also Baier, op. cit., Australasian Journal of Philosophy, 1962, pp. 57-68, and Smart's reply in the same journal, pp. 68-70.

contingent correlations between the subject's and the computer's accounts which are buttressed by surrounding theory and successful applications of theory become quite invulnerable to scattered cases of what Quine calls recalcitrant experiences.[1] The burden of explaining the occurrence of the single recalcitrant result remains, but the procedure is to attempt to explain it within the established theory and not to throw out the theory and start from scratch.

Even if no recalcitrant results are recorded for a theory, it is not possible to prove that no recalcitrant results could occur; a man could walk through a stone wall some day, and if this happened the physical sciences would no doubt be thrown into turmoil. Until such time, however, as such feats are well confirmed, the best confirmed tenets of physics stand as fact, not as mere conjecture.

The effect of this argument is to change the stance that must be taken by anyone who wants to call mental events necessarily private. Analysis may produce the conclusion that according to our ordinary concepts mental events are private, but one cannot then argue that since mental events are "conceptually" private, there are necessarily private mental

---

1 "Two Dogmas of Empiricism", Philosophical Review, 1951, reprinted with revisions in Quine, From a Logical Point of View, 1953. See also Putnam, in Hook, op. cit., pp. 164, 168-71.

events, for a method has been proposed for abandoning ordinary language and concepts in this instance. It is left open for the believer in mental privacy to argue that the method described would never succeed, that neurology will never develop to such an exalted state, but that is an argument to be supported by physiological evidence and not by any supposedly a priori considerations. The rejection of the subject veto could only be defended by the establishment of a working physical theory which could describe the physical forces operating, and provide a near-perfect basis for prediction. The denial of this possibility cannot now be supported except by arguments about the limitations, intrinsic or extrinsic, of the physical capacities of the brain or of the hypothetical mind-reading apparatus. The operation of the subject veto in ordinary language is sensible and useful, since it accounts for the contingent facts about our present lack of knowledge about the wellsprings of human behaviour, but its peculiarity of function cannot be deified into a necessary truth about these wellsprings, without a considerable argument not just about our natural concepts.

Thus Ayer's view that we are logically bound to the subject veto is too strong if he means necessarily bound. Ayer has failed to see that there are conditions under which the subject veto could be legitimately denied. Perhaps this is because he supposes that anyone claiming to read minds would be some kind of self-styled clairvoyant or telepathist, and not a scientist. Ayer is right of course about the subordination of telepathists' reports to the reports of the subject. The telepathist is not, at least right now, a scientist. The scientist, with a wealth of supporting data and a credible physical theory to accompany his predictive power, could have the right to deny the subject veto - in the name of science, if you like - while the telepathist with only predictive power has no such right. Denying the subject veto for reports of mental experience may at present be a deviant use of language, but as Putnam argues, it is only contingently deviant, and our understanding of ordinary language may evolve as a result of scientific discovery so that such usage loses its deviancy.[1]

---

1 Putnam, in Hook, <u>op. cit.</u>, p. 172.

CHAPTER 4.

The Function of the Brain.

7. Facts.

The rapid growth of neurological knowledge in the last decade has produced a fluid situation in which theories and interpretations rise and fall, in which researchers are in disagreement about what they are looking *for*, and visiting theorists from other disciplines contribute valuable insight and misleading suggestions in about equal measure. Neurologists have been enticed into answering philosophical questions (naïvely), and mathematicians have constructed complicated analogues for cerebral functions drastically underdetermined by hard data. My aim in this chapter is to present first the best-established and most crucial facts, and then the best-grounded theoretical contributions; in ensuing chapters I will deal with the philosophical implications of these minimal, firm scientific findings.

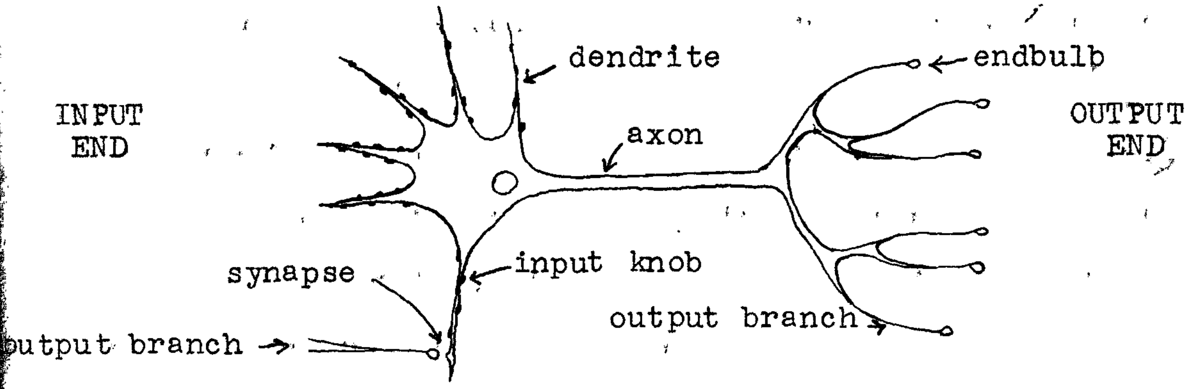
The major afferent or input pathway of the central nervous system, the dorsal tract, leads all its fibres to the cortex of the brain, where each input fibre branches out over a small area. This forms what has been called the man pattern on the cortex. Each area of a man's sensitive body is connected neurally to a particular area on the cortex, but needless to say, although there is more or less a point-to-point correspondence, the man pattern is not at all man-shaped.

When light strikes the retina, or when a point on a finger is touched by an object, "signals" are sent from the receptor cells on the retina or under the skin through the nerve network of the dorsal tract to the brain. The signals are in the form of repeating electrical impulses, travelling at speeds up to 200 miles an hour, all the same size but increasing in frequency when the stimulus is made more intense.

The signals of nerve fibres leading from the eye or ear do not differ in any apparent respect from other afferent signals. And the frequency of the impulses generally bears no relation to the frequency, for example, of the light or sound waves that stimulate the signals from eye and ear. There is nothing peculiarly visual or auditory, high or low pitched, about any signals arriving at the brain.

The networks of the central nervous system are made up of cells called neurones, which have an input end and an output end. The body of a neurone branches into dendrites, which are covered with hundreds of input knobs, and leading to each of these is an output branch from another neurone. The neurone's single output, or axon, branches after leaving the cell body into many outputs, leading to input knobs on other neurones.

(schematic drawing of a neurone).



The endbulbs of the axon branches do not quite touch the input knobs of the next neurones. The gap between them, or synapse, is crossed only when the frequency of the axonal impulses reaches a certain minimum level. Some synaptic crossings contribute to the excitation of the neurone and some inhibit its excitation. Each neurone has a "statistical" or "threshold" mechanism so that it fires its output only when the weight of excitatory crossings at a given moment exceeds the weight of inhibitory crossings by a certain value. To simplify, if each excitatory crossing is given a weight of +1 and each inhibitory crossing a weight of -1, a neurone with an excitation threshold of 3 would fire its output only when, at a given moment, the sum of all crossings is ≥ 3. In addition to these mechanisms, output branches from different neurones may intercept each other, before arriving at synapses, in such a way that simultaneous impulses in both branches cancel each other out.

Frequent firing of a neurone tends to lower its threshold, while inactivity raises the threshold. The explanation of this important phenomenon is not yet known. There is also evidence that particular synapses that regularly contribute to successful firings of a neurone tend to lower their frequency requirements, thus facilitating the crossing of impulses.[1] It has been observed that there is greater variation of diameter among frequently excited axons than among inactive axons, and since the diameter of the fibre determines the speed of the impulses carried by the fibre, and since nearly simultaneous arrival of inputs is crucial to neuronal firing, this variation in transmission speed can contribute to variations in neuronal firings. The net effect of these phenomena is to provide plasticity in neural networks; the functional structure of the networks is changed by patterns of stimulation, and these changes in structure determine changes in future patterns of neural activity.

---

1 It has been suggested that particular dendrites and end-bulbs grow toward each other, narrowing the synaptic gap, when crossings regularly contribute to the firing of a neurone, but this is unconfirmed. Sir John Eccles believes the synapse is crossed by a minute squirt of liquid. "The Synapse", Scientific American, Jan., 1965.

In addition to these details of individual cell function, certain gross phenomena have been observed to be necessary conditions for normal perception. The labyrinth or inner ear emits a barrage of signals varying with one's balance or position relative to gravity. If these signals are cut off or interfered with, disorientation can be so great that one is unable to recognise objects and coordination is severely curtailed.[1] Signals from joints relating to the position of limbs, neck and head also seem to be required for normal perception.

Another gross phenomenon is the widespread facilitation and inhibition of signals that occurs during concentration of one of the senses. During visual concentration, for example, there are "descending effects": a general facilitation or boosting of signals from the retina (through increasing the excitatory impulses to neurones along the visual paths), and general inhibition of signals from the ears, tactile receptors, and so forth.

---

1 Brain, "Space and Sense Data", British Journal of Philosophy of Science, 1960, p. 188. The new field of space medicine is investigating the possibility that the paucity of labyrinth signals during periods of prolonged weightlessness may cause widespread degeneration of the ability to think or concentrate. Thus on one view labyrinth signals are seen as "energizers" for the brain.

These facts, and a few more which will be mentioned in the next section, provide the foundation for a general interpretation of the function of the brain. The more detailed, and thus perhaps more satisfying, mathematical models of the brain that have been developed will not be considered in detail below because they are not yet secure.

8. Interpretations.

The chains of neurones, from the receptors right into the depths of the brain, are arranged in levels, so that the output branches of neurones in one level spread to reach dendrites in the next level. This directionality is not perfect, as is shown by the descending effects and other phenomena, but as a general rule - particularly in the cortex and at the retina - one can isolate levels of neurones at which the signals converging at the dendrites are at an equal remove from the receptors, and these levels in fact usually form a neat stratum identifiable by the anatomist. This notion of levels, which is systematized and heavily relied upon by many mathematical models[1] will be used more as an aid to brevity and clarity in this general account of neural functioning.

1 See, e.g., Arbib, Brains, Machines and Mathematics, 1964, pp. 5-7.

The function of the primary levels of neurones at the retina and the ear reveals a fundamental point about the character of neural signals in general. When a retinal receptor is stimulated, it fires impulses,[1] and these impulses can be called a signal, to the effect that light-waves within a certain frequency range have struck the receptor. But of course there is nothing about the signal itself that carries this particular content. The next neurones in line, to which this signal is transmitted, do not "read" the signal as information about light waves. They do not read the signal at all. The signal of that particular receptor normally fires if and only if such light waves have stimulated the receptor, and it is in this sense that the impulses can be given content by the neurologist.[2] Thus the impulses transmitted by any neurone or receptor can be seen as signals carrying a jot of information, this jot of information being determined[3] by the conditions of the signal's firing.[4]

---

1 I am simplifying; by receptor, I mean receptor proper plus internuncial cell.
2 The notion of endowing neural signals with content will be treated more fully in Chapter 6.
3 Determined in part. See Chapter 6.
4 I would prefer to use the term "bit of information" but this handy term has already been given a technical use in information theory (by J.W. Tukey), where "bit" is a contraction of "binary digit". I am avoiding the shorter alternative, "datum" since that term has unfortunately acquired an indestructible coating of bad theory at the hands of philosophers.

The mosaic of stimulation on the retina is not transmitted intact (via signals "about" points of light on the retina) to the brain. Right on the retina the first level of neurones breaks the information down into diverse jots of information about edges, motion, dimming, and so forth. The dendrites of the first level neurones branch in a number of different patterns, so that different types of stimulatory patterns on the retina cause these neurones to fire. For example, if a particular dendritic branching led to a vertical grouping of receptors, and the excitatory and inhibitory synapses and output interceptions were so arranged that the neurone would fire only when, say, the left hand side of the receptor cluster was stimulated, the information transmitted by this neurone would be about a vertical, light-dark boundary on that particular area of the retina. The information transformations on the human retina have not yet been analyzed, but in experimental animals, particularly the frog and the cat, signal "translation" up to the sixth and seventh levels of neurones has been achieved.[1]

---

1 Hubel and Wiesel, "Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex", Journal of Physiology, 1962, pp. 106-54; Lettvin, Maturana, McCulloch and Pitts, "What the Frog's Eye Tells the Frog's Brain", in Proc. of the Institute of Radio Engineers, 1959, pp. 1940-51, and "Two Remarks on the Visual System of the Frog" in Rosenblith (ed.), Sensory Communication, 1961; Muntz, "Vision in Frogs", Scientific American, March 1964, and Hubel, "The Visual Cortex of the Brain", Scientific American, Nov. 1963.

The effect of each level is to analyze further the reports of the lower neurones so that, for example, a high-level neurone fires if and only if a small, dark, moving object is in the visual field. Thus the <u>pictorial</u> representation of reality on the retina is broken down immediately at the retina and replaced with <u>reportorial</u> information. The information is interpretative in the sense that it is transmitted under certain specific conditions, and other conditions that may be evident in the pictorial representation on the retina of which the signal is a report are not adduced by the signal. Just as the description of the Mona Lisa, "a smiling woman with her hands in her lap", does not go into or adduce such evident conditions as colour, background, dress, or facial shape, so the reports of visual neurones carry limited, interpreted information.[1]

The stimulation of the ear is similarly broken down into discrete reports of types of stimulation. The continuum of a swoop on a slide trombone is broken into steps; receptors in the ear are primarily sensitive to one short range of frequency of sound waves, and the signals they transmit vary in frequency relative to intensity or volume of the stimulation, not relative to pitch (except at very low

---

1 Reportorial information will be discussed in detail in Chapter 7.

frequencies). But of course there is nothing to prevent a higher level neurone in the ear from being sensitive to swoops. It could fire if and only if a certain line or group of receptors fired in quick succession; differences in the time taken for the receptors' signals to reach the neurone's dendrites could allow the signals from the string of receptors to arrive simultaneously at the neurone, causing it to fire. Translation of higher level auditory signals has not yet been achieved, but there is no doubt that they are interpretative in the same way visual signals are.

These characteristics of neural activity have been embodied, with varying faithfulness to the data, in numerous analogues or models of the brain. W.S. McCulloch and W.H. Pitts pioneered these efforts with their concept of the logical neurone or module.[1] The McCulloch-Pitts neurone, a "black box" analogue, is an element that has the neural characteristics of threshold, multiple input, and single, branching output. These modules or neurones can be linked together to form "neural nets", the activities of which are uniquely determined by the initial input or stimulation. But neural nets made up of McCulloch-Pitts neurones are

---

1 McCulloch and Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity", Bulletin of Mathematical Biophysics, 1943, pp. 115-33.

simplified to the point that they do not accurately reflect the activities of the nervous system. First, all modules in neural nets must be considered to operate on the same time scale, with all synaptic crossings of a certain remove from the receptors occurring synchronously. Second, impulses are treated as binary digits, either on or off for each operational moment, whereas the impulses of individual neurones in the nervous system can vary independently in frequency. Third, the thresholds of the modules and the weights of the synapses are fixed. Many variations on the modules and nets have been proposed, some including threshold change, but all idealize neural activity to some extent.[1]

Instead of adjudicating between theories of choosing one model as the basis for further discussion, I wish to abstract three fundamental, shared features of all the models, features known to exist in the brain, and leave the details of their operation unaccounted for. First, it is clear that brains are learning organs, that stimulation causes changes in the functions of the brain. This is to be accounted for by changes in the functional structure of the

---

1 See, George, The Brain as a Computer, 1961, for a survey of these models and theories. Also, Konorsky, Conditioned Reflexes and Neuron Organization, 1948; Young, A Model of the Brain, 1964; Arbib, op. cit. The literature on the subject is enormous; George and Arbib provide good introductory bibliographies.

brain. Such changes (in threshold, for example) have been discovered in the brain, so the notion of storage through functional change is on firm ground. MacKay provides a good general definition of storage, applicable to computers and brains alike: storage is "any modification of state due to information received and capable of influencing later activity, for however short a time."[1] Second, these changes are brought about by the convergence of groups of input signals at individual neurones. Whatever the finer details of these changes - either physiological or mathematical - there can be no doubt that functional change depends on the firing of the neurone, and the firing depends on the convergence of inputs. Third, the information transmitted by the firing of a neurone depends on the conditions of its firing, which in turn may depend in part on its previous history of firing, if this has caused a functional change in the neurone.

Two obvious related points about neural functioning emerge from this very general account. First, the brain is "blind" to the conditions that stimulate its receptors. There is nothing intrinsically blue-reporting about signals triggered by blue light, nothing "C-sharpish" about signals triggered by

---

1 MacKay, "Toward an Information-Flow Model of Human Behaviour", British Journal of Psych., 1956.

the sound of C-sharp, nothing "big-toe-telling" about signals stimulated in the big toe. The brain performs operations with the signals that are triggered, but these operations are formal only, and depend not on what is "known" by the brain to have caused the signals, but on the convergences of signals that happen to occur. A computer solves problems without "knowing" that the signals fed into it refer in some way to missile trajectories, taxpayers, or votes cast, and the brain must be similarly blind to stimulus conditions. It makes no sense to suppose that the brain or a neurone could know or see what was stimulating it. If, in principle, the auditory and visual nerves arriving at the brain could be interchanged in the infant (a practical impossibility), he would not "hear" a symphony of colours and shapes, nor would he see sounds. The auditory and visual parts of the brain would simply switch functions and the child would see colours and shapes and hear sounds. This point is so fundamental and so truistic that it is difficult to give its expression any bite, but failure to see this point has led to some bizarre theories.

The second point is that the initial, unmodified interconnections of neurones in the infant brain can be random to some extent. The signal of a neurone acquires its functional significance for the brain through the convergences that happen to occur at its dendrites. Different "wirings" of neural networks would produce different informational functions for the individual neurones, but the end product of the network, behaviour, would remain the same. No particular convergences of particular signals need occur, for example, for the brain to interpret the retinal mosaic; whatever convergences do occur will suffice to differentiate the patterns of stimulation. The patterns of neural activity triggered by the projection of a uniform letter "A" on the retinas of different people can be as different as fingerprints. Provided only that the patterns in one brain for different stimuli are different, any sufficiently directional and richly interconnected wiring of neurones will produce the same discrimination.[1] This point, and the point about the formality or "blindness" of the brain's operation are often smothered by neurologists. Thus J.Z. Young says "we have to learn to see the world as we do" and Brain suggests that the baby must learn "the sensation language". These remarks suggest that there is a right way for the brain to organize its stimulus information; but any

1 Cf. the philosophical saw: if I see everything green as red, and _vice versa_, then I will have learned to use "red" and "green" for green and red respectively, and thus my descriptions and discriminations will match everyone else's. See Chapter 9.

way the brain organizes itself is "right" provided the behaviour it controls is successfully discriminatory. Thus to use Brain's metaphor, the baby, or more precisely the baby's brain, must make a sensation language, not learn one; any language will do.[1] Of course in fact there is a considerable degree of apparently inherited structure in the human nervous system, just as there is in the retinal connections in the frog and the cat, but there is no reason for this inherited structure to extend throughout the brain. Chapter 5 will deal with the question of the minimum structure that must be inherited if a brain is to be "successful".

9. Complications.

This account of neural function must now be complicated. Up to now there has been the tacit assumption that the firing of one neurone is determined by one stimulus condition, that the content of any particular signal is, however, difficult to determine, unambiguous. But this is not so. Even at low levels of neuronal interaction a wide variety of different stimulus patterns may cause the same neurone to fire, so that its signal is highly ambiguous. This fact,

1 See Arbib, op. cit., Ch. 2, on "Structure and Randomness", in which he discusses the Perceptron group at Cornell, which has simulated "randomly" wired, learning "perceivers".

distressing to the neurologists intent on "breaking the neural code", is, however, vital to the successful functioning of the brain. The brain, for all its occasional lapses, is a highly reliable organ; seldom if ever does a complete failure of stimulus interpretation occur. If each neurone had only one function, and this function was not duplicated by other neurones, the death or malfunction of any neurone would throw all that followed it into disorder. At the peripheral level - near the retina, for example - the death of a cell might only cause a small "blind spot" or imperceptible loss in colour discrimination or something of the sort, but if a single neurone at a high level is to carry single-handedly some bit of information about highly complex patterns of stimulation, its breakdown would cause something like total blindness for particular objects or wildly mistaken identification of objects in the visual field. Neurones do not regenerate like other cells, and their mortality rate may be in the neighbourhood of one neurone a minute. Neurologists have estimated that random malfunction of about one per cent of the neurones in the operation of any structure is normal.[1] Clearly, the reliability of the brain is greater than that of its components.

---

1 Arbib, op. cit., p. 54.

Arbib presents a calculation to show the effect of random failures: "Consider a chain of n modules [neurones] and assume that there is a probability p of malfunction for each neurone. Then the probability that the output of the chain is correct is, to a first estimate, $(1 - p)^n$. Now no matter how small p is, $(1 - p)^n$ gets to a 1/2 when n is made large enough and if our output is equally likely to be right or wrong, it is of no use to us!"[1]

Reliability of transmission using unreliable elements can be achieved, provided there is signal duplication in some form. If, for example, a message is transmitted simultaneously by five neurones, and the probability of successful transmission for each neurone is high, say .99, the probability that successful transmission will occur in at least three channels is much higher. Then, if a statistical or vote-taking mechanism is inserted between each level, random errors will be absorbed as soon as they occur. The variable threshold mechanism in the dendrites could perform this function, provided the redundancy of signals is great enough, and there is a rich enough interconnection of outputs with next-level neurones.

---

1 Arbib, op. cit., p. 56. Arbib points out that if p is one per cent, a neural chain of only 70 elements will have a probability of correctness of 1/2, and 70 elements is not very deep for the human brain, which has over 10^{10} elements.

But simple redundancy, with each neurone's signal serving one purpose, would require an inefficient multiplication of elements. If, on the other hand, the signals fired by each neurone are ambiguous (as they are), if each neurone contributes to many different multiple transmissions, redundancy can be achieved with less elements. It is then the more or less simultaneous concatenations or patterns of signals that are unambiguous, rather than the individual signals. The convergence of different concatenations of ambiguous signals at each succeeding level would partly resolve the ambiguity just as the "convergence" of ambiguous definitions determines unique or practically unique solutions to crossword puzzles. But there is no theoretical reason why the resolution of ambiguity should be completed before the final output signals of the brain, the last rank of efferent neurones that stimulate muscular contraction. Last rank signals trigger motions and hence are _de facto_ unambiguous. Malfunction in the last rank cannot be absorbed, but this just means that what is done is done. If there is a system of feedback, last rank malfunctions can be recognized as malfunctions, and "disavowed" or succeeded by corrections, but, of course, mistakes in the corrected last rank functions cannot be absorbed. A "slip of the tongue", once triggered, cannot be recalled, but it can be corrected; the correction, of course, may also contain a "slip of the tongue".

Information theory provides the thesis that reliable transmission with unreliable elements if possible, and many models of reliable channels have been made and put to use in the construction of reliable computers.[1] These provide hints about how the brain ensures the reliability of its operation, but as yet the details of the brain's particular solutions of this problem are unknown. Certainly the redundancy and ambiguity of signals contributes to the reliability. The spread of transmitted information over relatively wide areas may also increase the versatility of the brain by allowing subtle shadings of "emphasis" and partial facilitation and inhibition of parts of the concatenations of signals.

Widespread death or malfunction of neurones can cause noticeable breakdowns in behaviour; the brain cannot absorb

---

1 See Edwards, <u>Information Transmission</u>, 1964, and Arbib, <u>op. cit.</u>, for surveys of Shannon's theory, the Cowan-Winograd theory, von Neumann's multiplexing scheme, etc. Discussion by information theorists about coding and decoding has misled some neurologists into supposing that somewhere in the brain (or the mind) the signals would have to be "decoded" into a representation of reality. Neural signals may be transformed or <u>recoded</u>, but their "messages" need never be <u>decoded</u> into their original form (onto an inner retina of some kind, for example). See Brain, "Some Reflections on Brain and Mind", in <u>Brain</u>, 1963, p. 389.

serious damage any more than the reliable computer can absorb the blows of an axe. But the brain can often overcome severe damage, provided the damage is localized. If a particular area of the brain is injured, it cannot be healed, but another part of the brain can sometimes be made to take on the functions of the injured part. The functional structure of the brain is not entirely rigid, and perhaps not only because of neuronal losses and the resulting adoption of new functions. Neurologists have mapped the "man pattern" on the cortex of living brains only to discover that a few months later different connections existed, that cortical areas first excited by stimulation of one part of the body later were excited by stimulation of a different part of the body.

10. The General Picture.

The picture of neural functioning that emerges from all this is not heartening for those who would "translate" brain processes in the manner described in Chapter 2. Compared with the neat mathematical models and actual hardware computers, which are themselves awesomely sophisticated constructions, the brain's functional structure is vertiginously complicated. Instead of neat, digitally coded signals advancing in phase through a geometrically limned network there are signals varying in frequency advancing at different rates of speed through a shifting tangle of connections. There is high-order ambiguity of individual signals, and apparently high order redundancy and overlap of triggering conditions. The brain, with 10,000 million neurones (and even more glial cells, which may play a part in permanent or temporary functional changes), is many times larger than any computers now in existence. To further complicate the problem, the action of the reticular system in producing waves of increased sensitivity on other parts of the brain adds yet another layer of variables to the already highly variegated environment of stimulation in which we live.

So it is obvious that the type of highly predictive application of theory envisaged in Chapter 3 will remain a practical impossibility for myriads of man-hours (and computer-hours) of research and theorizing to come. Discounting truth-serums and other such short-cuts, our thoughts will be private for decades. But out of all this complexity of data and theory a few notions can be gleaned from which safe general observations on the nature of cerebral activity can be projected. Since all the details are not forthcoming, the projection of general guidelines should proceed now.

We can talk in a general way, with a high degree of faithfulness to the known details and the possibilities of the unknown details, about the content of signal-concatenations, this content being determined (in part) by the conditions of firing of the concatenations-as-wholes. (There are difficulties over the expression of content determined by conditions, and these will be dealt with in Chapter 6.) And content grows or is filled out at succeeding levels in the neural networks by the multiple convergences of the signal-concatenations. That is, the function of converging signals at individual neurones is spread through the concatenations, so that it is possible to speak of the convergence of concatenations-as-wholes. Storage of particular jots of information (which should be seen now as being carried by concatenations rather than by individual signals alone) is achieved by the functional changes produced by the carrying concatenations, and these changes affect the histories of future concatenations. Facilitation and inhibition, phenomena observed in the activity of individual neurones, can also be seen as spread over the concatenations.

Since I shall not be discussing the detailed operations of individual neurones from now on, I shall redefine "neural signal" to refer to the concatenation of firings that are produced at one level by a stimulus. The point is to reserve the term "signal" for the activities that can be safely described as having a certain content. This course allows the following description: the major functions of the central nervous system are the production of signals with certain contents, the convergence of signals to produce signals of enlarged (combined) content, the storage of information through structural changes produced by signals, and the facilitation and inhibition of signals through various types of signal interaction. This general description is vastly oversimplified, but, I hold, basically sound. Wherever pressures on these descriptions of functions occur, recourse can be made to the finer details. These terms provide a foundation for the general description of the neural activity involved in the varieties of mental experience (holding to the line that whenever one normally talks as if there were a mental experience, there is nervous activity).

It may be recognized that my account of neural functioning is incomplete in that I have said next to nothing about the efferent or motor side of the nervous system. The familiar notion of feedback, so common in discussions about the brain, has been mentioned only in passing. This is because this admittedly central notion requires additional theoretical foundations. Feedback must be negative or positive; the brain must interpret feedback signals as encouraging or discouraging (in a stretched sense), and yet I have said that signals in themselves are not intrinsically anything of the sort. There is nothing intrinsically painful, for example, about the signals triggered by injuries or painful stimulations, and yet these signals cause the organism to avoid certain courses.[1] In general, for the brain to develop the useful functions it does, there must be some mode of built-in encouragement of useful functions and discouragement of futile or harmful functions. Machines with servo-mechanisms are designed to have certain fixed purposes, and the servo-mechanisms are <u>designed</u> to guide the machines away from self-defeating courses; how does a similar functional recognition of "unhappy" information get established in the brain?

---

1 There are nerve cells that transmit nothing but pain-signals, but there is nothing intrinsically painful (whatever that might mean) in the impulses transmitted.

CHAPTER 5.

Pain, Evolution, and Feedback.

11. The Birth of Pain.

Every useful brain must come equipped with some partial means of discriminating stimuli appropriately. The development of the brain's functions depends on the sorting out of functional structures, keeping the useful ones and eliminating the harmful and useless ones, and unless there is some built-in standard in the brain, the intrinsically neutral signals of the afferent nervous system can never be endowed with significance in determining control of the organism's behaviour. The newborn child has a stock of built-in-reflexes and these inherited reflexes, I shall argue, are not only immediately crucial in insuring its nourishment and protection from harm during the first few years of life, but are the necessary foundation for all future development of the brain. The establishment of reflexes and the concomitant biological endorsement of certain types of neural activity can be explained in evolutionary terms.

At some point in evolution organisms appeared with simple nervous systems; contact with their surfaces produced electrical activity similar to that of neurones. The value of this phenomenon depended on the result it happened to trigger. Imagine three different strains of a certain type of primitive organism in which a certain stimulation or contact caused different "behaviour". In strain A the stimulation happened to cause the organism to contract or back off; in strain B the only behaviour caused by the electrical activity in it was a slight shiver or wriggling; in strain C the stimulation caused the organism to move towards or tend to surround or engulf the point of contact causing the stimulation. Now if the stimulation in question happened to be caused more often than not by something injurious to the organism, strain A would survive, strain B would tend to die off, and strain C would be quickly extinguished (all other conditions being equal). But if the stimulus happened to be caused more often than not by something beneficial to the organism, such as food, the opposite would happen. Then, although all three responses to the stimulation are blind, the response that <u>happens</u> to be appropriate is endorsed through the survival of the species that has this response built in. This observation is perhaps tautological: what is appropriate tends to aid survival; what is inappropriate tends to kill off the organism. The species that survive are the species that happen to have efferent signals connected to the incoming afferent signals in ways that help them survive.

As the evolutionary process continues, the organisms that survive will be those that happen to react differently to different stimuli - to discriminate, in other words. Thus if strain A backs off for both stimuli, x and y, while strain B backs off for x and increases contact for y, and if x happens to be an injurious stimulus and y happens to be caused by nourishment, strain A will die of starvation since it runs from both danger and food, while strain B will survive by discriminating. The discriminatory behaviour of strain B is only blind or dumb-luck behaviour, the appropriateness of which is revealed by the survival of the strain.

Once a variety of afferent-efferent connections has been genetically established, the various types of afferent signals involved acquire an inherent significance, as stimuli-to-withdraw-from and stimuli-to-remain-in-contact-with, and evolution ensures that the former are, in fact, danger signals and the latter are beneficence or security signals. In this way nervous systems evolved that are useful to organisms in prompting them to avoid harm and seek good.

Some stimuli are produced by injurious events or things and others are produced by beneficial events or things, and as a matter of fact we are prompted by the former to withdraw and by the latter to press on. This fact has been explained here without recourse to special and mysterious qualities of the stimuli that are somehow recognized by the organism or its brain or mind as criteria or indicators of appropriate action. Evolution can ensure that appropriate action is taken "automatically", as a result of built-in connections between afferent and efferent structures. The alternative explanation of this phenomenon which is inherent in our ordinary talk of pains is, in fact, no explanation at all. We say ordinarily that we withdraw from certain stimuli "because they are painful" - as if this advanced the explanation of how we happen to do the appropriate thing. But consider what could be the next step in the explanation. The initial question is: How do we happen to avoid harmful things? If the answer is that harmful things cause painful sensations, the next question is: How are painful sensations discriminated from pleasant sensations? The answer cannot be because painful sensations hurt, and pleasant sensations do not, since that leads to the question: How do we discriminate sensations that hurt from sensations that do not? - and a regress sets in.[1] Even if it is decided that pain and pleasure are

---

1 Of course a neuronal impulse could not hurt a neurone (although it might damage it); neurones do not have nervous systems.

unanalyzable qualities of sensations, there remains the question: Why do we avoid pains and seek pleasures? That is, even if we were to grant that the brain might distinguish these unanalyzable qualities of sensations, why should one quality lead to withdrawal, the other to advance? The move from injurious events to painful stimuli does not advance the explanation; it postpones it. One cannot build a bridge between sensation and action with qualities like "pain".

The solution lies in recognising that painful stimuli are just those stimuli that cause an organism to avoid the stimuli or withdraw. The statement that organisms avoid pain and seek pleasure is not contingently true, but tautologous. Pain only appears on the evolutionary front when organisms start avoiding it; pleasure only appears when organisms start seeking it. Neurologists have discovered no intrinsic characteristics of neural signals by which these signals could be discriminated by the brain as pain or pleasure signals; now it is clear that even if there were such qualitative differences, there would still be the problem of how and why the brain reacted as it did to these differences, which is really no different from the original problem: how and why the brain controls appropriate behaviour when stimulated by injurious and beneficial events. When it is recognized that the responses are mechanically coupled to the stimuli, and the appropriateness of the coupling is ensured by evolution, the question of recognized qualities in virtue of which the responses follow lapses. Similarly, one does not ask by what occult qualities a lamp switch determines that it is "on".

Of course "pain" and "pleasure" defined in this way diverge considerably from their ordinary language counterparts. For example, some people seek pains and shun pleasures, ordinarily speaking. Such discrepancies as are important will be mentioned later; the present appropriation of the two terms for theoretical purposes follows a tradition of using "pain" and "pleasure" as tag-names for certain opposing forces, qualities, or other theoretical entities (as in Utilitarian credit-debit schemes for determining the moral value of actions or Freud's "pleasure principle"). For the present I shall not talk about pains and pleasures but about pain signals and pleasure signals, leaving the question of whether there are pains and pleasures in the ordinary sense, and if so, what they are, for later.

12. Evolution in the Brain.

The genetic branding of pain and pleasure signals (and their necessarily accompanying responses, which characterize them as pain and pleasure signals) on the organization of the nervous system provides a basis for further functional development if there is some kind of association process in the brain. Since the first experiments by Pavlov, there has been considerable speculation about the fine details of the association process and rather than enter into the neuro-mathematical fray I shall again shun particular models and simply provide a broad description of the association process that few if any specialists would disavow.

If the occurrence of a pain signal is regularly accompanied or closely preceded by some other particular signal, this accompanying or preceding signal can become associated somehow with the firing of the withdrawal-triggering signal or signals so that the occurrence of the accompanying signal <u>without</u> the pain signal suffices to trigger withdrawal. This description says little more than is said by the obvious description of the external conditions of the association phenomenon in laboratory animals: if any neutral (intrinsically meaningless) stimulus is made regularly to accompany or closely precede a painful stimulus, the animal will eventually respond to the neutral stimulus even if the painful stimulus is omitted. The association function has simply been carried into the brain, where its machinery no doubt resides.

As Charles Taylor points out, there has been considerable theoretical difficulty in explaining how to "get to" the efferent response from the stimulus.[1] Skinner has the concept of "operant" behaviour, which is not stimulated or "elicited", but just "emitted" by the brain.[2] But operant behaviour, according to Skinner, can be connected to a stimulus cue somehow. If it is supposed that operant behaviour, such as the apparently random babbling of infants, is in fact not just emitted but stimulated (inappropriately - for no purpose) by the still unstructured afferents, the problem is not how the afferents can somehow make their way to the efferents, but just how the appropriate connections are weeded out. The difficulty for the psychologists seems to be that they are conceiving of the brain as initially divided into two spheres, afferent and efferent, that must somehow be bridged; somehow the stimuli must get to and connect with the appropriate responses. But if the brain is seen as initially richly (but "meaninglessly") interconnected between afferent and efferent, the task is one of sorting and short-cutting, and the background image of an afferent wandering through the brain in search of an appropriate efferent to stimulate dissolves. So does the image

1 The Explanation of Behaviour, 1964, pp. 117-20.
2 Science and Human Behaviour, 1953.

of a satisfied efferent in search of the right, significant afferent.

The association process allows a hierarchy of afferent-efferent connections to become established. Through association, injurious stimuli cannot only be withdrawn from, but avoided entirely. The initial pain-signals depended for their evolutionary establishment on being caused by injurious if not quite fatal impingements on the organism; through the association process, signals can trigger withdrawal by merely being (extrinsically) <u>harbingers</u> of injurious impingements. That is, there is no inherent meaning that makes these signals harbingers; it is simply their regular relation to pain-signals (which is determined by the way things are in the environment) and the <u>fait accompli</u> of their triggering of withdrawal which makes them act as harbingers. Other signals can come to be harbingers of harbingers and so on, and with similar growth among pleasure signals and associated signals, the organism will develop more intelligent, more cautious behaviour. Some of the second-order connections might in time become hereditary, probably depending on the regularity and uniformity of the harbinger signals. It would seem to make sense that the more tenuous the relationship between the occurrence of the pain signal and the occurrence of the harbinger signals, the less chance the second-order organization has of becoming hereditary. Blinking when something swiftly approaches our eyes is a good candidate for a hereditary second-order organization. The best established harbingers would be practically indistinguishable from pain signals.[1]

Different from withdrawal but related to it is inhibition of actions. When a pain signal or harbinger signal triggers withdrawal, the firing of the withdrawal signals must block out any opposing efferent signals; if efferent signals are firing causing the organism to approach or engulf a stimulus-causing thing, and pain signals occur prompting withdrawal, the approach signals (which may be "random" or produced by preceding pleasure signals) must be abrogated or overruled for withdrawal to occur. The survival of any species would depend on this hegemony for pain signals.

The result of the hegemony is the establishment of an evolutionary situation within the nervous system itself. The occurrence of pain signals or higher order negative

---

1 The view that mechanisms that are acquired in one animal might come to be transmitted genetically in another does not require a Lamarckian treatment. The hypothesis that a mechanism acquired by one animal can then be transmitted (under certain conditions) to its offspring is a much stronger view - but not yet a totally discredited one. Changes in RNA in trained animals at least suggest that an open mind is in order.

harbingers would inhibit the firing of inappropriate efferents, and any fortuitous appropriate efferents that occurred would not be inhibited and their occurrence, would increase the likelihood of their recurrence (because firing tends to lower neuronal thresholds and hence increase the likelihood of recurrence). The more useful an efferent happened to be to the completion of the triggered response, the more entrenched it would become in the brain. But entirely useless though entirely harmless efferents might also become established. A possible example is the tail-wagging of a happy dog. By association, these opposing forces of encouragement and discouragement would work their way up into higher and higher levels of neural organization, aligning the intrinsically neutral signals into appropriate organizations. Conflicts between pain and pleasure signals would sort themselves out by selective reinforcement and inhibition. For example, the beneficial effect of consuming nourishment is offset by the detrimental effect of "over-eating". The opposition of the resulting pleasure and pain signals would result in the inhibition of the consumption efferents when a certain point had been reached, thus ensuring an Aristotelian moderation in all pleasures. Conflicts between attack and retreat, action and inaction, sinking or swimming would similarly iron themselves out - not just in the long run history of the species, but in the short run biography of the individual organism. Thus the genetically endorsed pleasure signals and pain signals take over the evolutionary rôles of the carrot and the stick, and brains can develop appropriate behaviour controls within themselves and during their own lifetimes.

The advent of distance receptors such as eyes and ears greatly enriches the afferent barrage of useful stimulation, but allied with the development of distance receptors must be the development of more sophisticated motor controls. Efferent hierarchies must be established to govern motor activity sufficiently complicated and enduring to carry the animal to or from or around the things in its newly enlarged horizons. Advanced eyes would be no use to an animal without the capacity for controlled long-distance locomotion, and hence could not be produced independently by evolution. Increases in sensory input and improvements in behavioural output are interdependent.

Animals with distance receptors can perform appropriate behaviour over greater territory and longer time spans. They can learn, for example, that certain stimulus patterns, followed by locomotion in the direction of the stimulus-causer (followed by efferent activity that changes the stimulus patterns in the way they would be changed by the animal's approach to the object) lead to pain signals, and in this way they learn to avoid approaching the source of these particular stimulus patterns. The learned behaviour is selective since if the animal shunned every stimulus it would soon die of hunger. Caution must be balanced by action. The key to successful use of distance receptors such as the eye is the animal's "discovery" (effected by the reinforcement and extinction of chance efferent activity along the optic nerve) that certain patterns of stimulation should be facilitated and that certain controls of eye motion are more successful than others. The animal learns, one might say, to "pay attention"[1] to useful patterns of stimulation. In this way the learning process doubles back on itself, so that part of the behaviour controlled with the aid of the increased stimulation is behaviour designed further to improve and increase useful stimulation, which in turn serves to control still more sophisticated behaviour.

Under evolution, complexity begets complexity. Breakthroughs in behavioural deviousness must be met by more deviousness in those whose environment is altered by the new behaviour. Failure to adapt brings extinction. But the adaptation need not be in behaviour control; it may be in

---

1 This is a very figurative, non-ordinary use of the idiom "to pay attention". See Chapter 8.

physical prowess, protective colouration, poisoned spines, and so forth. In man the balance happened to tip towards deviousness and the ability to learn faster than other animals, and once the trend was started, the advanced nervous system developed at the expense of other adaptive possibilities. The day has passed when the species might have developed wings.

It should be noted that all complex animal behaviour (with the temporary exception of verbal behaviour, which will be discussed in Chapter 7) can be explained as controlled by the mechanisms suggested <u>without a concept of consciousness</u> or anything like consciousness coming into the picture. The introduction of a partial system of inherited and appropriate afferent-efferent connections into the brain establishes within the brain a means of organizing all afferent and efferent signals so that highly sophisticated and appropriate behaviour results, without there being anything intrinsically meaningful in any signal, without there being an internal eye that perceives or recognizes that particular signals report the existence of particular qualities or things in the outside world. The relationship between neural signals and their stimulus conditions or their triggered actions is "unknown" to the brain - how could it be otherwise? - and yet the <u>brain</u>, without the help of a knowing mind of soul or ego, can direct the activities of the animal with remarkable discrimination and success. Must a missile-guiding computer "know" or be "aware" that it is guiding a missile to its target if it is to operate successfully? It operates successfully because built into it are means of reacting appropriately to feedback, or in other words, discriminating between discouraging and encouraging signals. It performs its functions blindly and automatically; it performs them well if it is well designed. Evolution of the species ensures that brains are well-designed in part; evolution within the brain itself ensures that further functional structures are well-designed. Appropriate actions can then occur without there needing to be a "consciousness" of the action or a "consciousness" of facts about the external world in the light of which the action is appropriate. That is, there need be nothing over and above the means of signal organization described here; if that organization or some aspect of it is considered to be consciousness, all well and good, but there is no need to posit anything beyond that organization or any emergent phenomena produced by that organization.

Charles Taylor, in The Explanation of Behaviour, says:

> Systems to whom action can be attributed have a special status, in that they are considered loci of responsibility, centres from which behaviour is directed. The notion 'centre'

> seems very strongly rooted in our ordinary view of such systems, and it gives rise to a deep-seated and pervasive metaphor, that of the 'inside' ...
>
> What is essential to this notion of an 'inside', however, is the notion of consciousness in the sense of intentionality. To speak of an 'intentional description' of something is to speak not just of any description which this thing bears, but of the description which it bears for a certain person, the description under which it is subsumed by him. Now the notion of an action as directed behaviour involves that of an intentional description. For an essential element involved in the classification of an action as an action of a certain type, i.e., as directed to a certain goal, is the goal to which it is directed by the agent, i.e., the description it has _qua_ action for the agent ... Thus 'action' is a notion involving that of intentionality, _and the types of systems to which action can be attributed are those to which consciousness or intentionality can be attributed, beings of which we can say that things have a certain nature or description 'for' them._[1]

Now certainly nervous systems in living animals are centres from which behaviour is directed, but it is also clear that things do not have a certain nature or description "for" them, that they are not conscious of things as certain things or under certain descriptions. Then are the motions that (non-speaking)[2] animals go through really

---

1 _Op. cit._, pp. 57-8. My italics.

2 I am not denying that speaking humans can be explained along the same lines as non-speaking animals; I am not now positing some special characteristics for humans. I am simply postponing the discussion of human consciousness until Chapter 8.

actions, since Taylor insists that to call something an action is to give it the description it has 'for' the agent? One might hold that animal motions are not actions, but this is not a very helpful capitulation; animal motions are certainly just like actions if not indistinguishable from actions. We certainly <u>want</u> to call them actions. Are we then bound to posit some emergent consciousness in the animals?

Leaving the detailed answer to this question for Chapter 8, some small step in the answer can be taken now. Earlier Taylor says: "To offer a teleological explanation of some event or class of events, e.g., the behaviour of some being, is, then, to account for it by laws in terms of which an event's occurring is held to be dependent on that event's being required for some end."[1] But one should not extrapolate from this: "<u>known</u> to be required for some end by the agent". It is the <u>motion</u> that, under the circumstances, happens to be appropriate or required; the giving of a name or description of the motion as an action <u>interprets</u> the motion, explains its appropriateness. But the agent need not be conscious of or cognizant of the interpretation of the motion in order to produce it at appropriate times. Tauto-

---

1 <u>Op. cit.</u>, p. 9.

logically, it is only the interpreter or explainer who needs to interpret the motion, and agents, at least in the sense of producers of sophisticated motion, need not be interpreters of their motions. Again, the computer need not be conscious of the description or interpretation of its activity in order to generate prime numbers. Its activity is appropriate because it produces the motions it was programmed to produce; it is only the understanding or using of its motions that depends on interpreting them as generating prime numbers. Hilary Putnam points out that the answer to the question, How does the computer work out the decimal expansion of π? has three distinguishable constituents: "(i) A description of the sequence of states through which [the computer] passed ... (ii) A description of the rules under which [it] operated ... (iii) An explanation of the rationale of the entire procedure."[1] Without (iii) the computer's operation is incomprehensible; but describing the rationale is not positing any extra features of the computer's operation - least of all consciousness. It might be argued that the computer's makers or programmers must be conscious of the description of the action, but then who or what is conscious of the description of animals' actions? The brain-computers in

1 Putnam, "Minds and Machines", in Hook (ed.), Dimensions of Mind, 1960, p. 154.

animals are the product of evolution, the product of the world environment. Must we posit consciousness of some entity identical to the world throughout evolutionary history?[1] In fact, just such an idea lurks behind evolutionary theory; how strange and wonderful that apparently purposive, teleological structures and events are sifted out of the long history of interaction between organisms and environment! What are eyes for? For seeing. What made them with this noble purpose in mind? They just evolved. When it is grasped that teleological structures can evolve in an essentially non-teleological environment, the temptation to posit consciousness of purpose disappears; the temptation to posit an entity in which consciousness of purpose is to reside disappears. But slowly.

13. How Pains Hurt.

There is more to be said about what I have called pain and pleasure signals and their derivative, hierarchical counterparts. It is tempting to describe the basic, inherited pain signals as "The ones that really hurt", and the derivative signals as "merely unpleasant" or "mildly disagreeable", and to treat the fundamental pleasure signals as

1 Brigid Brophy, writing of Jane Austen, says: "She is an 18th-century writer in the sense that she is the novelist that century had been trying to produce all along." New Statesman, 4 Dec., 1964, p. 879.

"acute pleasures" or "visceral pleasures", treating the afferent barrage that occurs when one reads poetry or listens to Buxtehude as "intellectual pleasures" or in any case, as tepid (however sublime) imitations of animal ecstasy. But obviously, talk of pain signals and pleasure signals is not of the same type as talk about things as unpleasant, disagreeable, mildly amusing, acute, and so forth. The two types cannot be so simply mixed. The situation is better described in more neutral terms, pain-signals and their derivatives being called "negative" signals and pleasure-signals and their derivatives, "positive" signals. Then one can point out that the line between basic negative signals, which are produced by actually injurious conditions, and derivative negative signals, which are mere harbingers of injury, cannot be easily drawn in practice. The differentiation of degrees of negative signals depends not so much on their firing conditions as on their weight in controlling behaviour. If one were to divide negative signals into pain signals and, say, mere displeasure signals, the division that would fit most closely with the ordinary concept of pain would be on the basis of how "compelling" the signals were. In general, the farther removed a negative signal is from the inherited foundation of negative signals, the more easily it can be overruled by a marshalling of positive signals. For example, the positive weight of the olfactory afferents stimulated in a mouse by the presence of some food may be outbalanced by the negative weight of some visual afferents stimulated by the presence of a hawk, and thus the mouse withdraws. Talk in terms of the positive and negative weights of signals can replace, with negligible loss, talk in terms of the qualities and intensities of pains and pleasures. We say, ordinarily, "I had such a splitting headache I could not enjoy the concert", "The sublime taste of lobster is worth the bother of shelling it", and "The pleasure of winning a marathon is not worth the pain of running in one". It would be simple-minded, of course, to start searching for specific lobster-eating pleasure signals in the act of overruling shelling-bother pain signals, and that is just to say that it would be simple-minded to suppose that the particular entities proposed by our ordinary talk - in this case, pleasures and bothers and pains - have neat counterparts among the structures and events in the brain. The point is that the system of weights and balances that I have suggested as operating in the nervous system is adequately complex to handle, among its configurations, the differentiations suggested in ordinary talk. If we were in the habit of saying (and presumably meaning and understanding) statements like "Every time I reach for a lobster claw I can feel the pleasure of the lobster taste goading me on, doing battle with my feeling of displeasure at having to shell the claw" then it would make some sense to look for particular, relatively discrete patterns of neural firings that could be called "shelling-bother negatives" and so forth. But we do not speak in that way, so it is unlikely that any neat correlations could be found, and even if we did speak that way, no bald identity of brain processes and pleasures would be proclaimed.

But someone may object that pain is something extra, in addition to the signals that prompt withdrawal. One might be suddenly and overwhelmingly compelled to remove one's finger from a hot stove without the additional phenomenon of pain occurring. But although withdrawal is the basic function of pain signals, and is the only function in very low animals, such as slugs and worms, in higher animals the "compelled" behaviour is more complicated, and some of it is not obviously appropriate; dogs yelp and squirm and jump about and lick their wounds; people cry and say "Ouch" (if English) and moan and dance about clutching the injured area. How are these additional compelled actions to be explained? (By compelled, I mean that the transition from afferent to efferent is so strongly established and so heavily weighted that normally no other barrage of signals can overrule it.) Licking wounds no doubt is appropriate in helping to damp the negative signals, for as long as the negative signals occur the animal is bound to a course of avoidance, and in higher animals avoidance can become not merely a matter of contracting away from a stimulus. Dancing about may be some vestigial and inappropriate attempt at avoidance, a "running away" from the stimulus that does not happen to work. Other superficially inappropriate activity might have the effect of bombarding the afferent channels with non-negative signals which have the effect of partially inhibiting the basic negatives. In any case there is no guarantee that all the pain behaviour controlled by the brain is appropriate. So the objection becomes: one might, on hitting one's thumb with a hammer, be suddenly and overwhelmingly compelled to drop the hammer, suck the thumb, dance about, shriek curses, moan, cry, etc., and yet <u>still</u> not be experiencing pain. That is, one would not be acting, as on a stage; one would be compelled. One would be unable to respond to polite applause with a smiling bow. What sense can be made of saying that there is some strange quality or phenomenon beyond all this that is pain? The key to the "painfulness of pain" is the compulsion of the behaviour, something which is not externally evident in behaviour, and thus the behaviourist who will not look beneath the skin finds pain to be private. But it is only contingently private.

Wittgenstein says, at one point: "'Yes, but there is _something_ there all the same accompanying my cry of pain. And it is on account of that that I utter it. And this something is what is important - and frightful.'"[1] He puts the remark in quotes, so perhaps it is not his view; in any case its wording confuses him, for later he says: "'And yet you again and again reach the conclusion that the sensation itself is a _nothing_.' - Not at all. It is not a _something_, but not a _nothing_ either! The conclusion was only that a nothing would serve just as well as a something about which nothing could be said."[2] One wonders if his mother tongue has bewitched him here. He is looking for the difference between genuine and feigned pain-behaviour, and I can only conclude that if any sense at all can be made of his something-nothings, his position is compatible with mine, except that I say something about that "something about which nothing can be said". What is there in some slight displeasure beyond one's readiness to behave in ways so that the displeasure (the readiness to behave thus) disappears? All there is more than this is the permeating and confused

1 _Philosophical Investigations_, i 296.
2 _Ibid._, i 304.

belief that there must always be entities behind our ordinary way of speaking. We do not always treat "pain" as referring to an entity: "Why are you dancing about like that?" "Because I'm in pain." Surely that idiom can be explained without mentioning any object, pain, that one is in (as one can be in a box). Keats' pale knight was held "in thrall" but where is the temptation to look for thralls? Keats' idiom has disappeared in favour of the verb "enthrall"; could one not exclaim, with no loss of meaning, "Because I'm enpained"?

I have just said that there is no guarantee that all pain behaviour is appropriate (in the special sense of contributing to survival), and this holds for all behaviour. If mice made no mistakes, hawks would go hungry. The tendency of evolution, both of the species and of the individual brain, is to preserve the appropriate, and this tendency is stronger the closer the behaviour concerned is to the brute facts of survival. As behaviour becomes more devious, less directly related to survival, the corrective force of the stern environment is dissipated, and the question arises: Appropriate for what end? That is, appropriateness for survival becomes too broad a measure, and it is convenient to break it up into more manageable parts by speaking of appropriateness for some intermediate state of affairs, these states of affairs being generally conducive, however indirectly, to survival. But in man the complexity of behaviour has snowballed; behaviour appropriate in the individual is inappropriate from the point of view of the community or society. Survival is barely relevant to our mundane, daily activities, the appropriateness of which (for such ends as wealth, happiness, security, fame, education) is obvious and assured, while the imponderable choices set before political leaders, scientists, and even the voter, bear directly on survival. Man has developed behaviour controls that can range over years, by establishing hierarchies of behaviour as complex means to the attainment of distant and ephemeral pleasures, and the avoidance of distant and dubious pains. At such a stage the effect of evolution's corrective force is nil, and so it is not surprising that wildly inappropriate behaviour becomes established in certain individuals and societies. Perhaps the most directly inappropriate activity is committing suicide; next is masochism, in which the higher development of neural functions partially reverses the direction of the inherited functions on which it is based. But then the question can be asked: Inappropriate for what end?

A peculiarly human means of raising the complexity of behaviour is the introduction of systematized artificial pains and pleasures, used to _train_ other beings - humans and animals - to perform artificial responses. In this way we tame animals and teach the young to speak, from which the bulk of human deviousness derives.[1]

Of course the more abstruse needs, pleasures, and pains, such as the need for a graduate degree, the pleasure of seeing _Hamlet_, and the pain of boredom, are far removed from the basic "instinctual" mechanisms, and the devious gathering of information by medieval historians has only the slightest relevance to behaviour control. But the birth of these phenomena is dependent on the existence of their ancestors, the inherited pain and pleasure signals.

14. _Feedback_.

The simple notion of a negative feedback mechanism is one in which "signals", whether impulses or light waves or the motions of simple levers, cause a mechanism to respond in such a way that the cause of the signals is removed or diminished. The spinning governor on a steam engine is the classic example. Applied in its simplest form to an explanation of human behaviour control, e.g., reaching for a book, the notion yields a picture of the proprioceptive signals

---

1 See Chapter 7 on the inculcation of language.

from the joints and muscles combining with the visual feedback to guide the hand to the book by triggering, facilitating, and inhibiting the various efferent neurones that stimulate muscle contraction. And if this picture is true in general outline, the details present complications.

First, there are many controlled activities that occur at speeds too fast for simple stimulus-response feedback to influence. A pianist trilling on a note raises and lowers his fingers faster than signals can travel from fingertip to brain and back to finger muscles. Presumably, the pressure of the fingertip against the depressed key stops the firing of the efferent that pushes the finger down and triggers efferents that raise the finger and lower the other finger, but the interconnections that could do this could not occur as fast as trilling actually occurs. Evidence points to the cerebellum as the part of the brain responsible for producing alternative methods of high speed control. T.C. Ruch comments that "slowness of voluntary movement is characteristic of cerebellar patients and of normal individuals executing unpracticed movements."[1] The hypothesis is that for well-trained motions, such as typing, writing, and trilling, the

---

1 "Motor Systems" in Stevens, ed., Handbook of Experimental Psychology, New York, Wiley, 1951, p. 204, quoted in Miller, Galanter and Pribram, Plans and the Structure of Behavior, 1960, p. 91.

control is transferred to the cerebellum, where a considerably different kind of control mechanism takes over.

Miller, Galanter, and Pribram[1] present an informative model of such a mechanism, which is somewhat supported by neurological data. They call their model a Test-Operate-Test-Exit unit, or TOTE for short. The simple TOTE unit tests as it fires its efferent signal. Part of its efferent signal contributes not to muscular contraction but to the stimulation of stretch-sensitive receptors in the muscle; the afferent signals from these receptors return to the TOTE unit, where their frequency is "tested" against the frequency of the efferent signals. The TOTE continues to fire until discrepancy between the two disappears.

> The action is initiated by an "incongruity" between the state of the organism and the state being tested for, and the action persists until the incongruity ... is removed ... Thus there is "feedback" from the result of the action to the testing phase, and we are confronted by a recursive loop.[2]

By building behaviour control out of hierarchies of short feedback loops, instead of temporal chains of firings, each controlled in turn by one central headquarters, swifter and more complex behaviour can be controlled. Hammering in a nail can be seen as a two-level TOTE structure, the low level

---

1 Ibid.
2 Ibid., pp. 25-6.

TOTEs terminating their activity at the conclusion of each up or down stroke of the hammer, the high level TOTE terminating when the nail head is flush, that is, when certain visual feedback reaches the pre-determined cut-off level. The authors draw a distinction between reinforcing feedback and TOTE feedback.

> (1) A reinforcing feedback must strengthen something, whereas feedback in a TOTE if for the purpose of comparison and testing; (2) a reinforcing feedback is considered to be a stimulus, or information (e.g., knowledge of results), or control (e.g., instructions); and (3) a reinforcing feedback is frequently considered to be valuable, or "drive reducing", to the organism, whereas feedback in a TOTE has no such value.[1]

Thus TOTE feedback is neither positive nor negative, but merely incongruous or not. The authors seem to think that this is a "more general conception of feedback".[2] But it cannot be the <u>primitive</u> feedback in the brain. Somehow TOTEs must be established to perform appropriate functions; in machines this is the responsibility of the designer; in brains, of the pre-existing pain and pleasure signals. The mercenary TOTE can perform any function, and its allegiance to appropriate functions must rest on something. Reinforcing feedback is essential.

The concept of feedback in the nervous system, in whatever theoretical terms one decides on, is here to stay.

---

1 *Ibid.*, p. 30.
2 *Ibid.*, p. 31.

Research into feedback mechanisms has been fruitful, allowing precise models of systems which can control types of behaviour distinguishable in normal animals and humans. Further support is offered by the fact that the kinds of malfunctions exhibited by man-made servo-mechanisms (feedback mechanisms) are also evident in nervous disorders. Norbert Wiener describes two such disorders:

> A patient comes into a neurological clinic. He is not paralyzed, and he can move his legs when he receives the order. Nevertheless he suffers under a severe disability. He walks with a peculiar uncertain gait, with eyes downcast on the ground and on his legs. He starts each step with a kick, throwing each leg in succession in front of him. If blindfolded, he cannot stand up, and totters to the ground. What is the matter with him?
>
> Another patient comes in. While he sits at rest in his chair, there seems to be nothing wrong withbhim. However, offer him a cigarette, and he will swing his hand past it in trying to pick it up. This will be followed by an equally futile swing in the other direction, and this by still a third swing back, until his motion becomes nothing but a futile and violent oscillation. Give him a glass of water, and he will empty it in these swings before he is able to bring it to his mouth. What is the matter with him?
>
> Both of these patients are suffering from one form or another of what is known as ataxia. Their muscles are strong and healthy enough, but they are unable to organize their actions. The first patient suffers from tabes dorsalis ... The receptors in the joints and tendons and muscles and the soles of his feet, which ordinarily convey to him the position and state of motion of his legs, send no messages which his

central nervous system can pick up and transmit, and for information concerning his posture he is obliged to trust to his eyes and the balancing organs of his inner ear. In the jargon of the physiologist, he has lost an important part of his proprioceptive or kinesthetic sense.

The second patient has lost none of his proprioceptive sense. His injury is elsewhere, in the cerebellum, and he is suffering from what is known as a cerebellar tremor or purpose tremor. It seems likely that the cerebellum has some function of proportioning the muscular response to the proprioceptive input, and if this proportioning is disturbed, a tremor may be one of the results.

We thus see that for effective action on the outer world it is not only essential that we possess good effectors, but that the performance of these effectors be properly monitored back to the central nervous system, and that the readings of these monitors be properly combined with the other information coming in from the sense organs to produce a properly proportioned output to the effectors.[1]

Clearly the first patient suffers from a loss of feedback; the disorder in the second patient is strikingly similar to the behaviour of governed engines that "hunt". Often, due to some slight maladjustment of the levers of mechanical governors on engines, the speed of the engine will oscillate widely around the desired speed instead of gradually approaching stability. Such an engine is said to "hunt", and the cause of the hunting is over-compensation triggered by the governor.

---

1 Cybernetics, 2nd edition, 1961, p. 95.

To recapitulate the last two chapters, Chapter 4 described the way the brain can organize its input signals into useful patterns of high discrimination, without the signals having any intrinsic content or content-giving characteristics; Chapter 5 has shown how the corner can be turned from afferent signalspatterns to efferent signal patterns, without the invocation of a ghostly interpreter of signals' significances. This lays the foundation for physicalistic descriptions of awareness, intending, knowing, and perceiving, and for physicalistic answers to the questions that have been acknowledged and postponed in these chapters. The first of these that must be considered is the problem of assigning content to neural signal patterns and neural structures for the purposes of explanation.

CHAPTER 6.

Content and Function.

15. The Leap to Content.

In Chapter 3, I argued that afferent signals (in the wide sense of signal patterns, if not always in the narrow sense of individual neuronal firings) could be ascribed a certain content, depending on their stimulus conditions. A signal could be given the content that p if it fired (normally) if and only if p. The difficulties with that account, which was designed as a stopgap, must now be cleared up.

The move from description in terms of neural firings, neural structures, and the regularities observed in these, to description in terms of the content of signals and structures is a large theoretical leap. When one talks about content - and its companions, meaning, reference, intension, significance - one is in a different world of explanation (although I am not prepared to give a name to this world) from the world of explanation when one talks about chairs and tables, electrons and impulses, paper and pencil. One does not make the mistake of attributing meaning to the trails of ink on the page; chairs do not refer to anything; electrons have no significance. So, it seems, the neurologist who ascribes meaning or content to neural impulses must be involved in the same kind of ascent, vaguely, as is the critic or logician who discusses the meaning of written or spoken words. If this point be recognized, let me drop it; discussion of it in these woolly terms will only breed mystery. I wish to consider the case in more practical terms.

Suppose the psychophysiologist of the future has access to the neural activity of a dog, and he observes the dog refusing to venture out onto thin ice to retrieve a succulent steak placed there by the experimenter. The experimenter has the following information: high level visual signals, similar (in stimulus conditions and the path they take) to signals recorded on previous sensory confrontations with steaks, trigger salivation; they also trigger high level efferent activity normally present when the dog is about to approach or attack something, but the continuation of this efferent activity is blocked through inhibition by signals traced to structures that developed when the dog once fell through thin ice.

This fantasy is crude, and no claim is being made that this is just the way the future psychophysiologist would describe his findings, but if this were the gist of his findings, he would be in a strong position to explain the dog's behaviour as controlled or directed by the stored information that the ice was too thin to walk on. Such a description, that is, would be better than descriptions like "the dog did not notice the steak", "the dog has an aversion to smooth horizontal planes", "the dog is in a trance", "the dog is overcome by Weltschmerz". The visual signals would be given a content something like "there's a steak", the efferent signals the content "get the steak", the inhibitors the content "stop; it's not worth the discomfort", and the storage structure the content "if ice looks like x [which it does in this case] it is too thin to walk on". Now aside from the fact that the particular contents ascribed to these signals are imprecise, not strictly determined by any of the data, and overly fanciful, it is clear that they help explain just what is going on in the dog. But it is also clear that the psychophysiologist with practically complete information on the dog's neural history would not need to use these ascriptions to predict or explain the neural behaviour or the external motions. The dog's brain does not rely on the ascriptions of content; the ascriptions of content are not "for" the dog or his brain. The brain's activities are organized by physical events and conditions in the nervous system regardless of any content-giving interpretation, and thus they can be explained and understood within the world of physical explanation, just as events on a billiard table can be explained and understood without recourse to fanciful interpretations (such as "the white ball ordered the red ball to hit the cushion, and it obeyed"). Since the brain is a closed, "blind" system, the neurologist's account or mapping of the system could be similarly blind. The neurologist could label the neurones and signals by any system of numbers or letters that was convenient, and his theory would explain and predict on the physical level alone. But the neurologist could not explain the <u>appropriateness</u> of the activities he could predict. There are two questions with two different types of answer. First, "Why did efferent E fire?", which has an answer in terms of synapses, thresholds, and so forth; and second, "Why should such an appropriate controlling activity take place?", or "How is that activity appropriate?". The long answer to the second question involves a discussion of evolution, the environment that happens to be the case when certain nervous activity occurs, and the effect that the animal's motion has on this environment. By speaking in terms of content the neurologist can explain the appropriateness of the neural activity in the same breath as he describes its causal relations. Endowing neural signals with content is a short-cut way of keeping their appropriateness in mind, and thus an aid to deciphering the plainly appropriate activities of the brain.

When the original strain A backs away from an injurious stimulus, one would not say it withdrew from a painful stimulus. One would say it simply did what it did because of the physical connections that happened to exist, and as chance would have it, the motions were appropriate to the survival of the strain. But as natural selection gradually produces organisms that consistently move appropriately (their less fortunate cousins in strains B and C having died off) one can say that strain A did what it did because of the physical connections that happened to exist, and these connections happened to exist because they were appropriate - but not because any being recognized them as appropriate. At this point one can give particular types of appropriate responses to stimulation action-descriptions like "withdrawing from pain-stimuli" and "seeking out food", and these descriptions are teleological, but do not presuppose anything emergent (be it intuition of purposes, consciousness of pains, or whatever). They merely presuppose the long evolutionary story which could explain the survival and reinforcement of the particular controls.

Most of our ordinary description of action is interpretative. The physical, "motion-only" description of signing a contract gives no clue to its appropriateness.[1] The description, "signing a contract" has a dual rôle; it refers quite neatly to the physical motion and simultaneously interprets it, alludes to its appropriateness. Of course simply calling neural impulses signals is interpreting them in the light of their appropriateness. And, perhaps, talking of inhibition and triggering is similarly interpretative. But all this means is that one can hardly open one's mouth without "interpreting" the straight physical facts. The leap to content of neural signals is just a particularly straightforward leap. The theoretical leap to talk of content is gratuitous in principle in any purely physical explanation of the brain, but in practice the neurologist who denied himself the understanding provided by the interpretation would be hard put to devise hypotheses to be tested.

The decision to talk of content of signals, or of stored information (the content of structures or conditions), is like the decision to give an interpretation to a formal mathematical system. Discussing a formal system under an interpretation does not affect its structure or powers, but

1 In contrasting the world of "physical" explanation with the world of contentful or teleological explanation, I do not mean to suggest that the latter implies the existence of anything non-physical or emergent or irreducible. By "physical" I mean within the relatively non-interpretative vocabulary of the non-life sciences.

often makes it considerably easier to understand. In the case of neural systems it should be recognized that content is not some property or entity that signals or organizations or structures have in addition to their functional capacities, but that talk in terms of content is a useful interpretation of functional capacities. One establishes the interpretation by relating functions systematically to descriptions of sensory confrontations and muscular motions, framed in "appropriateness language".

16. Practical Difficulties.

The gift of content, however, is not so easily given as that. Ascribing content is giving the signal a message to carry, and this message must be expressed in words. How else, in a written or spoken account?[1] But which words? At what point, for example, do afferent signals carry messages about objects in the world, and not just about patterns of light on the retina, or pressure patterns on the skin? Perhaps the best point to ascend to object-talk is after visual and tactile signals have converged, but since the whole process of endowing signals with content is governed by pragmatic rather than strictly theoretical ends,

---

1 There is no escaping from this via *oratio obliqua*; "The signal signals *that* there is a red object in position a,b,c" is no better than "the signal's message is 'red object at a,b,c'".

it may be convenient to give pure visual signals (or even, in some animals, olfactory signals) "objective significance", the justification being the use to which the animals' brains put the signals. The point is not to find absolutely certain grounds for "belief"in the objective world", but to find the "grounds" which in fact suffice to control objective behaviour. No argument against solipsism is at issue here, since no animal behaves solipsistically - refrains from behaving, that is.

There is, however, a deeper seated difficulty over the choice of words to express the messages. Consider the case of the dog and the steak. If the neurologist has pinned down a high level visual signal (with perhaps some convergence of olfactory signals behind it) that for various reasons is decided to have a content something like "There's a steak", any English message the neurologist relates to this signal is bound to imply differentiations not revealed by the functional capacities of the signal. Should the signal be translated "steak", beefsteak", "meat", "food", tenderloin"? Supposedly the signal's functions reveal greater differentiation than the message "food" would indicate, since, for example, the dog is more interested in steak than dog biscuits, but the other possibilities seem to suggest too much. "Meat" suggests that the dog recognizes the food as butchered animal part, and how could any canine behaviour indicate such a discrimination? "Beefsteak" suggests the further recognition of a particular type of butchered part of a particular type of animal, which is even more absurd. Clearly, the differentiations in a dog's brain would not match those of English or any other human language - it would be astounding if they did. A zealously rigorous neurologist might go to the trouble to formulate the language of Doggish, the words of which were defined (in Doggish) to match the dog's differentiations. This course would be immensely tedious, and translation of Doggish into English would be necessarily imprecise, since if it could be precise there would be no need to formulate Doggish. And the neurologist could not say that the dog thinks in Doggish, but just that he, the neurologist, describes the dog's thinking (the dog's cerebral processes) in Doggish, and what is the value in this? Since any disputes over differentiation would be settled by examining actual neural functions, by reverting to the world of physical explanation, approximations of content in ordinary language would provide the wanted explanatory power just as well as the more cumbersome, and only speciously more rigorous ascriptions in artificial language.

The less quixotic course would be to ascribe content in English with an eye to keeping the information of the ascribed messages to a minimum, and a caveat to the reader to avoid extrapolating with the aid of a dictionary. The sea lawyer who squeezes every possibility of discrimination out of the expressed message is as much of a pointless nuisance here as in everyday communication. For, do we not all use "steak" without qualms or misunderstanding, and yet how many of us could provide a strict definition of the word, or, more dramatically, cut a proper steak out of a side of beef? For most words, the differentiations of English speakers do not match the dictionary differentiations of English.

In §48, I shall show why these attachments of groups of words (in English or in Doggish) to neural signals must of necessity be a rough guide only. There is a fundamental difference between messages in words and neural "messages" which prevents any systematic ascription of content to verbal signals. But this fact will not at all rescind the permission to use content-talk as a loose short-cut description of neural function, provided that the privilege is not abused by extrapolation. To safeguard against this in the meantime, I shall make more explicit the physical conditions that determine content.

Content is to be determined in part by sensory confrontation, in part by exhibited functional capacities. The functions of any neural signal involve the functions of many others - tautologically - since the functions of a signal are the triggering, facilitating, and inhibiting of other signals. The function of a signal thus depends on what has gone before to establish the paths in which it, and all it triggers, operate; to say it in circular fashion, function depends on what information has already been stored. At the same time, whatever descriptions of the sensory confrontation may be given, the information carried by any receptor or neurone must be determined by that part of the stimulus conditions to which it is sensitive. The messagesthat can be carried by neural signals, like the messages we can send by telegraph, are about a very limited part of the stimulus conditions; they do not "go into all the details", and cannot do so. The question: What features of the situation does the telegram report? is in some ways like the question: To what features of the situation is this neuronal system "sensitive"? (See §48.)

It is in the nature of a signal or report that it is selective (see Chapter 9). If one lacks information on the stimulus sensitivity of a signal or of the stimulus origin of a neural structure, one cannot give the hypothetical evolutionary story explaining its production and survival and hence cannot ascribe content to it. But knowledge of the stimulus origin of a neural signal or structure is not enough. The dependence of content on the capacity to direct behaviour is particularly plain in the case of information storage structures. If the afferent barrage attendant on a dog's burning itself in a fire causes a structural change that determines a change in the functions of certain succeeding signals, the content of the information stored depends directly on what these functional changes produce. If these changes do not produce avoidance of fires, then regardless of the sensory confrontation that established the structure, there would be no reason to describe it as storing the information that fire hurts.

The appropriateness, if any, of structural changes can be alluded to in a breath by describing them as storing information. But there is no thing, no granule or inscription, that is the stored information; there is only the capacity for controlling appropriate functions, and this capacity is produced by earlier sensory confrontations.

An appreciation of these conditions underlying content ascription allows the explicator of neural activity to make use of content-talk with a fair measure of safety; and there remains one escape hatch: when imponderable difficulties in the world of content appear, the explicator is to descend again to the level of physical explanation. The heuristic value of this content-endowing interpretation of the physical operations in the brain increases, the more complex the particular operations are. This is particularly evident in the explanation of efferent organization. There is little if any heuristic value in giving content to the "last rank" motor impulses that stimulate muscle contraction. Giving an impulse the message (perhaps best in the imperative form) "Contract now, muscle!" does little to clarify what is going on, if only because the physical account is already clear enough. Why should one say that pushing the starter button orders a car engine to start? There are, however, neural conditions that cause much more complicated behaviour, that are not bound to one particular pattern of lower level efferent firings. When the appropriate behaviour is describable as, say, getting the piece of meat on the table, a dog will try a number of ways of getting it if at first he fails. What is appropriate is not usually a particular configuration of muscular contractions but a particular changing of the relationship between animal and environment, or a changing of parts of the environment. "It has been observed that once an animal has learned to perform a task (say to run a maze to a reward) it will perform that task with whatever means are available to it. If its legs are amputated, it will roll through the maze."[1] The clear way to comprehend the development and operation of such structures and signals in the organism as a whole is to grant them content: to speak of the orders that these signals carry. The raison d'être (quite literally) of such structures and signals is to effect particular appropriate changes in the environment, and giving them content is a direct way of alluding to the rôle-dependency of their occurrence or existence in the brain. The practice is well established for computers. Computers are fed streams of intrinsically meaningless binary digits; some groups of these are interpreted (by the computer's operators) as directions or orders ("add the number 101101 to the number at 11101 and move to 11011"), but the computer does not so interpret them; it does not say, "Ah, here is an order for a little addition", but simply does what it is designed to do when binary digits (as impulses or marks on a tape or whatever) in certain sequences are fed into it.

---

1 The experiment is Lashley's (Brain Mechanisms and Intelligence, 1929). This particularly brusque expression of the result is found in Rapoport, "Technological Models of the Nervous System", Methodos, 1955, pp. 131-46, reprinted in Sayre and Crosson (eds.), The Modeling of Mind, 1963.

What sense can be made of brain processes if they are not given content? This does not mean: if brain processes do not have content, we cannot make sense of them. Written words do not have content; when we give them content we make sense of them.

CHAPTER 7.

Language and Certainty.

17. Linguistic Behaviour.

I have postponed detailed discussion of linguistic behaviour until now for several reasons. Man is the speaking animal, and the advent of speech is also the advent of all the most human phenomena, a behavioural explosion on the evolutionary scene. The problems associated with linguistic behaviour are, in a way, of a different magnitude from the problems associated with non-linguistic behaviour. And because of this burgeoning complexity it is here that I expect the greatest resistance to physicalistic explanations. The subject requires special attention.

Most higher forms of life have some established quasi-linguistic, quasi-communicative behaviour: beavers slap their tails in warning, bees do their little dance, dogs bark and growl at each other, and dolphins - according to latest reports - have a wide range of organized noise-making behaviour that may well deserve the name language. The appropriateness of this behaviour is obvious, and there is no reason why its endorsement by evolution should differ radically from that of other behaviour. I do not want to argue that this behaviour is or is not really linguistic or really communicative; it is in any case a pale copy of sophisticated linguistic behaviour. For one thing, most of this quasi-linguistic behaviour is rigid and unadaptive. It seems to be directly and permanently tied to particular types of stimulation, occurring even when the communal situations in virtue of which it is appropriate (and to which it owes its existence) are not in effect. Evolution, in "roughing in" the proper stimulus conditions may not allow for all the possible exceptions that would vitiate the appropriateness of the response. Also, the behaviour is narrow in scope and versatility.

An hypothesis would be that the behaviour in question has been established by historical (as opposed to intracerebral) evolution and is transmitted genetically. This would explain its uniformity and failure to adapt to immediate situations, or, as some might put it, its <u>lack of rationality</u>. (This point will be discussed in Chapter 10.)

Some animal behaviour of the quasi-communicative variety may be learned, or established by "evolution" in the individual brains. Separating "instinctive" from "learned" behaviour is notoriously difficult, but if it could be determined, for example, that a great part of the dolphin's noise-making behaviour is developed after birth through training, the claim that dolphins have a language would be much more acceptable. Behaviour that is learned is more adaptable since it can be unlearned if it proves inappropriate, without waiting for generations of the species to change their ways. The greater the brain capacity and the greater the afferent barrage and the greater the complexity of behaviour control (all these being interdependent facets of one situation), the more appropriate in detail, the more "cognizant" of specific situations behaviour can be. Thus in animals with relatively large brains, vocal behaviour seems more communicative.

The linguistic behaviour of man is certainly appropriate behaviour. The question whether it is something radically more than just appropriate behaviour of the type so far described, whether it is an emergent phenomenon and not simply a sophistication of ordinary behaviour, can be answered best by treating linguistic behaviour as only very advanced behaviour but of the same type as other behaviour. Then the task is to show that all the generally supposed linguistic phenomena and their peculiar characteristics can be explained (or explained away) in the course of description. The survey of these phenomena is the task of the rest of the thesis as a whole; the promise is that not only the supposed phenomena, but the traditional difficulties with these phenomena will be explained. The task of this section is to present a preliminary hypothesis of the dynamics involved in the production and control of linguistic behaviour.

It may be useful in passing to mention the presumed genesis of language. If the quasi-historical myth begins with a small range of grunts and sounds of approximately dolphin standard used to various ends such as warning, threatening, calling, and directing, the finer features of language can be seen to follow along certain general lines. The use of proper names for calling would perhaps be among the early bits of behaviour to become impressed on the primal horde as useful, and it is just a short step from this to the naming of objects. Importantly, there is no need to imagine anything like deliberate coinage of words at this stage. Possibly in the beginning all sounds divided roughly into a few groups, any guttural sound, say, meaning danger and any moaning sound being used to call; differentiations within the groups would gradually get impressed on usage by the action of fortuitous successes. Since any recurring afferent or efferent pattern could, in principle, come to stimulate a bit of this new behaviour, and since certain types of patterns predominate in their relevance to the survival of the group (those produced by familiar objects, qualities, motions, etc.), it would be natural for the differentiations of the linguistic behaviour to become tied to these patterns. The convergence of two or more of these could lead well enough to the concatenation of sounds, making primitive sentences, and so forth. A logical feature such as "not" would most likely develop from a previously used utterance "No" used in the prohibitive sense and gradually inserted in longer concatenations. The advent of a logical vocabulary would probably coincide with the advent of verbal responses to verbal responses.

Since many of the higher level efferent patterns of importance to the person's survival are not linked exclusively to particular motions or sequences of motions but to achieving certain appropriate changes in the resultant afferent barrage, if these patterns came regularly to trigger particular utterances, the utterances could come naturally to be interpretative in the way "signing a contract" is interpretative. The utterances could be interpretative because the conditions of their utterance would "exist for a purpose", and thus the utterances would always occur in conjunction with the occurrence of a teleological function. Hence the "gratuitous leap" of Chapter 6 is one to be expected in the natural development of ordinary language. Having a "way of describing actions with an eye to why they are done" (having a concept of action, in other words) is thus not a strange theoretical accrual to primitive language, but a feature to be expected of language.

At the same time as straightforward linguistic behaviour was becoming established, devious linguistic behaviour would become established. The appropriateness of occasionally keeping mum instead of speaking, and of occasionally speaking the wrong words would not long be lost on these founding fathers.

An important point connected with this story is that once animals (or humans) start living together in families or larger groups, there is a new vehicle for the transmission of appropriate modes of behaviour and hence for the evolution of higher modes of behaviour in animals with high capacities for individual adaptation. If training and imitation are established as appropriate modes of behaviour the experience of the old generation can be passed on to the young, and this training need not be anything as deliberate as schooling. The evidence seems to show that in many higher animals even such vitally important behaviour as mating is at least partly learned by the young, and is not inherited as brute instinct. The evolution of language, as studied by the philologists, is of course the prime example of this vehicle of transmission and evolution. Once communication has been established, of course, learning proceeds at breakneck speed.

So much for pseudo-anthropological daydreams. The details of the story are only fanciful surmises, and it was included only because if one wished to deny that truly linguistic behaviour could develop from the evolutionary swamps, one would have to deny the possibility of such a story being sound in general outline. There is still much to be said about linguistic behaviour in its particularly devious aspects, but these matters will be dealt with in detail one at a time.

The neurological picture of lingustic behaviour that follows from this leaves no gaps between the afferent barrage and the utterance of words. Afferent signals are shunted through the various control organizations established by previous neural activity and eventually signals arrive at that part of the brain known as the speech centre. These signals trigger signals that trigger signals and so forth until the last rank is reached; the motor signals fire that move lips, larynx, tongue, and lungs; an utterance is made.

Consider the content of these signals, starting from the last rank and working back. The last rank signals hardly need be given content. Their messages amount to "tongue forward", "engage voice box", "exhale", and so forth. But what of the content of the signals that organize these? Their content would be phonemic: "utter: 'la'". At the next level signals would fire that produce sequences of phonemes. Since spoken sentences are not strictly just phonemic sequences ("She can't bare her face" and "She can't bear her face" have the same phonemic sequences) it would be overstepping to give these signals a content like "utter: 'the cat is on the mat'"; it should rather be "utter: ... " followed by a phonemic rendering. But what of the next level and the next?

In some cases of linguistic behaviour the motions that will suffice to effect the appropriate change in the environment are strictly limited to the production of a certain phonemic sequence. When one takes an oath, recites poetry, or in other ways quotes someone or some document, the controlling efferent hierarchy's job is to produce the phonemic sequence (plus emphasis, tone of voice, etc.). A non-linguistic example of this bond between a particular sequence of motions and appropriateness is choreographic reproduction in ballet. In these cases the activity of the controlling hierarchies above the level of phonemic organization could well be given the *oratio recta* content of the uttered words: "say: 'I do solemnly swear ... " At this point one might ask the embarrassing question à la Ryle: how many same-level controlling hierarchies are involved in the recitation of "The Charge of the Light Brigade"? But in this case the question is not really embarrassing. Overriding the entire activity of reciting there must be some continuing temporary controlling neural state with the content, vaguely, "Recite what's put before you". This state, of course, can be overruled by various stimuli - heckling, cries of "Fire!", and so forth - but given comfortable conditions it is sustained until the task is completed and it guides the hierarchies beneath it which will be broken up differently for different speakers. The young reader may go word by word; the experienced reciter's hierarchies may govern the recitation of whole verses and possibly even stanzas. The embarrassing question has, in principle, a numerical answer, however difficult it might be to determine it in the individual case.

But in other cases, indeed in most cases, of linguistic behaviour, what is appropriate is not the production of a particular phonemic sequence so much as the effecting of a particular change in other people's behaviour or behavioural capacities (see §49), which may be done in a number of ways. The controlling hierarchy in these cases does not have an *oratio recta* content so much as an *oratio obliqua* content; it is analogous to the efferent organization in the rat that will carry it through the maze *in one way or another*. In assigning content to these hierarchies one runs into the practical difficulties described in Chapter 6. The scope of variation that is to be allowed to the ascription depends on the richness and versatility of the organizations that follow or are beneath the signals, something that cannot be alluded to by any one particular ascription. One might assign as content the alternation of all the different expressions in the repertoire of the sub-hierarchies that could be triggered by the controlling hierarchy (the content: "'p' or 'q' or 'r' or 's' ... "), but such an unwieldy expression would vitiate the purpose of assigning content in the first place, as an easily comprehensible, pragmatic short-cut. It is tempting to invoke the somewhat disreputable philosophical word "proposition" in this circumstance, but the difficulty here, as ordinarily with talk of propositions, is that they allow too much. Having come this far, I am going to back out of the difficulty; the solution to the problem of ascription in these cases is not to be met head on. The function and appropriateness of these hierarchies is clear enough, and if the short-cut of content ascription runs into snags, that is a linguistic matter. (See §48.) Provided one treads with caution, and keeps an eye on the foundation of function and evolution-endorsed appropriateness, talk of content can be used as an approximation. For example, a signal might be assigned the content:

"apologize!", and the lower levels of the hierarchy might effect this with "Pardon" or "Terribly sorry" or "Excuse me". It is tempting then to say that such a signal at this level is identical with the thought that accompanies one's apology, different signals occurring for the different thoughts (say, one such signal being the sincere thought that one is very sorry, another being the thought that protocol demands an apology), but there are difficulties, so for the moment I will leave it at that; it is tempting.

Now if this description in terms of controlling hierarchies is substantially correct, one would expect that there would be feedback loops associated with each level, that each level could be inhibited, and that there would be mistakes in linguistic behaviour to correspond to misfirings or regular shortcomings in each level. Stuttering would seem to be as close to a last rank malfunction as any common error is. Stuttering is certainly a mistake "closer to the muscles" than, say, calling John "Bill" by mistake. At the level of phonemic organization lisping and saying "tewwibwy sowwy" would be examples of malfunction or more often of permanently faulty programming. At the level of phonemic sequence there are spoonerisms: "Mardon me, Padame". The next level would be strictly <u>verbal</u> errors: "I'm terribly silly - I mean, sorry!". Also at this level there are malaprops that have been faultily programmed in the first place: "he writes all his books under a pomme de terre", habitual grammatical mistakes, and the like. Beyond this the mistakes are not linguistic mistakes, but bits of linguistic behaviour that are inappropriate to the situation however well-produced: "Oops, Dammit!" instead of "Please excuse me". One has said the wrong thing, but the mistake is not really a linguistic mistake any more than a *non sequitur* in a book is a typographical mistake.

The next question concerns what happens when each level is inhibited. If the signals triggering and directing linguistic behaviour are inhibited at any point on the way to muscle contraction no observable linguistic behaviour will occur, so there cannot be any overt phenomena to point to as involving inhibition at one level or another. But what about "hidden" phenomena?

If I may be allowed for the moment to step into the forbidden world of introspection, several suggestive points can be made. Sometimes when we think our thoughts we are "talking to ourselves" in a very strong sense; what we think is expressed in sentences, in definite words in a definite order, even in a particular "tone of voice" with particular emphases, and the thinking of these thoughts takes just about as long as saying the words would take. But sometimes our thoughts are not like this; sometimes they are swift, somehow not quite formulated into particular words, and not in a particular tone of voice. One might describe the former by saying that it was just like talking except that no muscular motions took place; one can talk, or one can whisper, or one can just move one's lips; or, one can eliminate even the lip movements and what is left is this type of thinking.[1] Then one might describe the latter by saying that what one eliminates beyond all this is the formulation of temporal sequences of particular words. Something of temporal succession remains, but it is not the same as the easily clockable sequence of words in the former case. It would be a very likely supposition that when one was talking to oneself the situation differed from when one was talking out loud in that the last rank efferent signals were inhibited, and that when the efferents were inhibited at the level just above phonemic sequence (or slightly higher) thinking without talking to oneself would occur. I am not saying that thoughts are identical with the signals just above the level of phonemic sequence. Such a move would be ill-judged considering the very figurative description of thinking that

1 It has been thought by psychologists in the past that this phenomenon involved incipient but not quite detectable muscular motions, but his view is no longer prominent.

has been proposed. Why tie quite confirmable physical hypotheses to metaphors? No sooner would I identify these thoughts with these particular signals than the following objections could be raised: How could I be <u>aware</u> of an x-level speech centre efferent signal?, How could such a signal be witty? I know what I am thinking certainly, but I do not know which particular x-level efferents I am having - if such signals do occur. So I will for the moment draw the line at hypothesizing that signals of the type I have described occur when one is thinking in certain ways and not at other times, and that the content of these signals (in the sense of "content" I have outlined) bears a certain striking resemblance to what one might ordinarily or philosophically call the content of the thoughts one is having.[1]

18. <u>The Certainty of Certain Utterances</u>.

If it is supposed that there is some strong relationship between thoughts and these signals, the first difficulty to be met concerns certainty. There is a widely shared intuition or feeling or belief that there is some absolute

1 These suggestions begin to make some physical sense out of Ryle's observation that thinking is just talking without moving one's lips.

certainty involved somehow in our knowledge and reports of experience. The difficulty in accounting for this intuition has elevated the notion of perceptual or experiential certainty into a philosophical and psychological problem of great importance. I intend to show that the problem of certainty rests on two minor linguistic muddles and one major pseudo-problem.

The first linguistic muddle concerns authority and incorrigibility. Part of the aura of certainty of introspective reports resides in the subject veto condition of ordinary language. The philosophical argument runs: since first-person introspective reports are necessarily (logically) authoritative, they are necessarily incorrigible, and if something cannot in principle be corrected it can be considered to be certain since it is impossible to deny it. The conclusion seems to be a <u>non sequitur</u> since statements that are contingently incorrigible (for contingent lack of evidence) can certainly be false, and removing the possibility of correction (or confirmation) does not seem to be quite the same as ensuring the truth of the statement; what seems ensured could equally well be the essential <u>un</u>certainty of the statement. But this is neither here nor there, since there is a muddle in the first premise. To the extent that there is "logical" authority to introspective reports, it is a matter of linguistic heritage - the subject veto - and could be changed by linguistic evolution or reform. By starting from authority one mixes the facts of the case with the facts of how we talk about the case. There may be some necessary certainty in introspective reports, but this could not be proved by the existence of an assumption to that effect embedded in ordinary language.

The second linguistic muddle is pointed out by Wittgenstein. "Imagine someone saying, 'But I know how tall I am', and laying his hand on top of his head to prove it."[1] Obviously there is no assertion in this statement and act; it is not even ostension, but going through the moves of ostension. One cannot be wrong if one does not commit oneself to an assertion. Such committal need not be in words. If I think I see a lion and fire at it I commit myself just as much as if I said "I see a lion". But if I say only "I see what I see", and perhaps gesture in the direction of my eyes or brain, I make no assertion and hence, trivially, cannot be mistaken. This point was missed by the sense-datum theorists who pared down the "red, apple-shaped sense-datum" to "red, round sense-datum" to "ruddy, roundish somewhat" to "this, here, now". Having arrived at

1 Philosophical Investigations, 2nd edition, 1958, i 279.

an incorrigible statement, they had also arrived at a remark of no interest or import whatever.[1] Their purpose was to eliminate what one could be wrong about, and when they had done that there was nothing left to be right about. If there is any interesting or important certainty in introspective reports it is not to be located by this method.

The major pseudo-problem rests, strangely enough, on a fallacy that is so universally recognized that it hardly deserves mention. And yet in subtle guise it creates puzzlement where none ought to exist. It is the fallacy of the little man in the brain who looks at the movie screen and listens to the ears' radio and then makes infallible reports of his experiences. The fallacy is brought back into the fold in the positing of sense-data or similar go-between entities between the outside world and reports of perception. The idea of sense-data involves the idea of a secondary stage of perceiving, a perceiving within perceiving (or an intellectual <u>having</u> or intuiting within perceiving), and since a person, a human being, can wrongly identify what he sees (what thing is out there), there is the problem of how the whatever-it-is that perceives or sees or just has the sense-data infallibly identi-

1 Ayer makes this point in several places.

fies them (what things are in here). Even if the whatever-it-is is strongly identified as the person himself, the problem remains of how the person's perusal of these intermediate entities is infallible. Saying that the person does not peruse sense-data but simply intuits or has them is just pointless obfuscation in the face of an absurdity.

I have outlined the neural journey from sense organs to effectors, explaining the production of behaviour produced by and appropriate to the sensory input, and nowhere along the line did I come upon the production of sense-data or their perusal. In fact, the journey was in an important sense homogeneous; the important properties and functions of neurones at the retina do not differ from the properties and functions of neurones in the brain itself or neurones leading to muscles; there are no peaks, gaps, bridges, turning points, or qualitative changes. What geography there is in the brain is determined by functional areas, but the functions of the individual neurones are uniform throughout. No feature or functional area of the brain comes close to resembling the infallible homunculus at the controls. The key to the dissolution of the problem of certainty is the elimination of the middleman.

Hilary Putnam points out that a computer that is programmed to ascertain the 3000th digit in π can be said to pass through a certain sequence of states, A,B,C, etc., in the process of computation.

> Now let us suppose that someone voices the following objection: "In order to perform the computation just described, T must pass through states A,B,C, etc. But how can T ascertain that it is in states A,B,C, etc.?"
>
> It is clear that this is a silly objection. But what makes it silly? For one thing, the "logical description" (machine table) of the machine describes the states only in terms of their <u>relations</u> to each other and to what appears on the tape. The "physical realization" of the machine is immaterial, so long as there <u>are</u> distinct states A,B,C, etc., and they succeed each other as specified in the machine table. Thus one can answer a question such as "How does T ascertain that X?" (or "compute X", etc.) only in the sense of describing the <u>sequence of states</u> through which T must pass in ascertaining that X (computing X, etc.), the rules obeyed, etc. But there is no "sequence of states" through which T must pass to be in a single state!
>
> Indeed, suppose there were - suppose T could not <u>be</u> in state A without first <u>ascertaining</u> that it was in state A (by first passing through a sequence of other states). Clearly a vicious regress would be involved. And one "breaks" the regress simply by noting that the machine, in ascertaining the 3000th digit in π, <u>passes through</u> its states - but it need not in any significant sense "ascertain" that it is passing through them.[1]

Suppose T "ascertained" it was in state B; this could only mean that it behaved or operated <u>as if</u> it were in state

---

1 "Minds and Machines" in Hook (ed.), <u>Dimensions of Mind</u>, 1960, p. 154.

B, and if T does this it __is__ in state B. Possibly there has been a breakdown so that it __should__ be in state A, but if it ascertains that it is in state B (behaves as if it were in state B) it __is__ in state B. One cannot ask "How does it know?" here just as one cannot ask of the mathematician who says "Let x equal 7", "How does he know?".

Putnam notes that there is an analogous fallacy in epistemology. I may know there is a chair in the room "on the basis of sense experiences E_1, E_2, etc.", but it does not follow that I must know or ascertain that I am having these experiences. Putnam's word "experience" seems ill-advised; the point can be better put: my saying or acknowledging that there is a chair in the room (or some behaviour such as this) is based on or produced by (but not concluded from) the afferent impulses A_1, A_2, etc., but it does not follow from this that I, my soul, or my brain is intuiting, or cognizant of, or perusing either the afferent impulses or any ghostly parallels to them.

J.J.C. Smart approaches this idea with his concept of "topic-neutral" introspective reports.[1] He would have it that introspective reports are equivalent in purport to (though not translations of) statements of the form: "What

---

1 __Philosophy and Scientific Realism__, 1963, pp. 94-8.

is going on in me is like what is going on in me when ... " where the dots are filled in with descriptions like "my eyes are open, lighting is normal, and there is a yellowish-orange patch on the wall". The "what" is topic-neutral; it may refer to brain processes, spiritual processes, or whatever. Smart's point is that one need not know what exactly is going on to say that what is going on is like or unlike what goes on at other times. Smart says:

> Now it is tempting, when we think in a metaphysical and a priori way, to suppose that reports of similarities can be made only on a basis of the conscious apprehension of the features in respect of which these similarities subsist. But when we think objectively about the human being as a functioning mechanism this metaphysical supposition may come to seem unwarranted. It is surely more easy to construct a mechanism which will record (on a punched tape, for example) bare similarities in a class of stimuli than it is to construct a machine which will provide a report of the features in which the similarities consist. It therefore seems to me quite possible that we should be able to make reports to the effect that 'what is going on in me is like what goes on in me when ... ' without having any idea whatever of what in particular is going on in me (e.g. whether a brain process, a heart process, or a spiritual process).[1]

Smart is on the right track, but he has mixed mentalistic talk ("without having any idea ... ") with physicalistic talk (recording mechanisms) into an indigestible stew. In fact, his mode of expression seems designed to puzzle and antagonize

---

1 Ibid., p. 96.

the critic of physicalism. Smart describes the topic-neutrality of reports thus: "A dualist will think that what is going on in him when he reports an experience is in fact a non-physical process (though his report does not say that it is) ... and I think that it is a process in my brain."[1] What is this thinking that the dualist is doing? What does Smart's thinking the process is a brain process amount to? The suggestion is certainly that something or someone (the physical person? part of the brain? the ego?) is at least aware or cognizant that there is <u>some</u> process going on. Can it be that the brain or part of the brain is aware that a process is going on, but cannot tell just what kind of process? Certainly what Smart means is that when he has an experience (and experiences <u>are</u> brain processes for Smart) he knows or is aware that <u>some</u> process is going on, and he has discovered through science that in fact it is a brain process. How does he come to know that <u>any</u> process is going on? He does not similarly know that his digestive processes are going on. What is it about Smart's brain processes that allow them to make themselves known (but only in their occurrence, and their similarities and differences to others) to Smart? This is what happens when

---

1 <u>Ibid</u>., p. 95.

one baldly identifies thoughts and experiences with brain processes; mentalistic talk and physiological talk simply do not mix.

And aside from this, Smart misses Putnam's point about ascertaining. Smart gets the point that the computer (in this case, the brain) need not ascertain what its structural or physical states are, but he does not see that the brain also does not ascertain what logical states it is in. The picture of brain processes that emerges from Smart's account has the brain divided into two parts, the presenting part, and the ascertaining part. Just as one can make a mechanism "which will record ... bare similarities in a class of stimuli" Smart presumably pictures a mechanism in the brain used to ascertain the similarities of certain impulse patterns, and it is these impulse patterns that he identifies as thoughts or experiences. Now if there were such a stage in the brain (and there might have been) the question "How does the brain ascertain that experience X is occurring?" would be a perfectly good question. For a person to be able to say that there was a chair in the room, certain experiences (brain processes) would have to occur <u>and</u> the brain would have to go through certain states to determine that this was so. On Smart's view (and I admit I am extrapolating somewhat; Smart never really says just what sorts of brain processes he is identifying as thoughts) there is a break or bridge or turning point on the road from sense organ to utterance. It is located at that stage where some neural mechanism checks stimuli for similarities. The mechanism takes over the rôle of homunculus, the stimuli are the sense-data, and the pseudo-problem remains: how does the whatever-it-is (now identified as a bit of undescribed neural machinery) determine (infallibly) just which sense-data (now stimuli) are occurring? It helps, before talking of brain processes, to find out what sorts of brain processes there are.[1]

Suppose some scientists wished to construct a machine that duplicated all the crucial physical features of perception. Its sense organs could be television cameras (two, for binocular overlap), and as an improvement on human sight it could also have a radar to provide extra information about size and distance. The impulses from these sense organs could be recoded in any regular way to fit the input requirements of an immense neural net analyzer which would feed its output into a "speech centre" computer. The speech centre computer would be programmed to "translate" the output of the analyzer into, say, printed English

---

1 Cf. Ibid., p. 102-3.

messages, like "I see a man approaching".[1]

It might just be worth mentioning that there would be no need for television or radar screens in this machine. Setting up such screens and then monitoring them with more cameras would simply postpone the activity of the analyzer. Television and radar outputs, unlike the output of the eye, are in the form of single sequential streams, and it would no doubt be advisable to "spread" the sequence of impulses reporting each complete scanning of antenna or television camera image by time lags over a bank of inputs so that single scannings were fed into the analyzer simultaneously. But there would be no reason (save reasons of engineering design, perhaps) why this spreading should follow the spreading pattern of an electron gun in a television tube and reproduce the image on the bank of inputs. Since nothing will be looking at or photographing the individual arrays of impulses over the bank of inputs, there is no need for the pattern of inputs to produce any image or topological analogue of the sense organ's image. Provided the spreading system is stable, any spread could be used.

Once the speech centre computer had been properly programmed to produce the appropriate English sentences as

---

1 Such a machine could not now be built; the techniques, time, and money are just not available.

expressions of the analyzer's output, the whole machine could duplicate human reports of visual perception. I am leaving out such features of human perception as focusing and facilitation of particular patterns, and any speech centre approaching human capacity would be enormously sophisticated, but presumably these features could be built in, given the time and money. And, of course, it would be a waste of paper and electricity to have the speech centre working continuously; in practice its input would be inhibited and only put through on a spot check basis. A breathless and repetitive babbling from the machine would be of little use. Now once the machine is operational, in what ways could its reports be fallible?

First, it would be fairly easy to trick the machine. Presenting it with moving dummies could result in the report "I see a man approaching", or, for example, the television and radar outputs produced when a man was approaching could be recorded and then fed into the analyzer at some time when there was no man approaching. In these cases, the analyzer would issue the same output as for veridical "perception". Its output would be just <u>as if</u> there were a man approaching. Aside from trickery, there might be malfunctions in the sense organs or the neural net analyzer. This would be guarded against by redundancy measures, but malfunctions would still be possible. An uncorrected error in the analyzer would issue in a mistaken output.

This mistaken output would be expressed by the speech centre unless it too made a mistake. Feedback loops in the speech centre would be designed to correct malfunctions before the actual printing took place, or if the malfunctions took place in the last rank - the actual printing - they would be erased and corrected. In spite of these precautions, speech centre mistakes could occur if the feedback malfunctioned.

But if "verbal errors" are discounted or corrected, whatever output does enter the speech centre will be correctly expressed <u>relative to the rules of language programmed into the computer</u>. Disallowing "misuse of language" and "slips of the tongue" there is no room for mistakes to occur in the expression of the analyzer's output. That output, whether right or wrong relative to the actual sensory confrontation, <u>cannot help but be</u> correctly reported if feedback corrects all verbal errors. An uncorrected mistake occurring <u>before</u> the analyzer's output <u>changes</u> the output to be expressed; whatever output is to be expressed will be expressed.

The key word is "expressed". The speech centre computer does not <u>report</u> or <u>describe</u> the output of the analyzer since the output is not a representation of what is outside, but a report itself. The speech centre does not examine the output to determine what qualities it has, but simply produces English sentences as expressions of the output.

There are two kinds of errors that the machine can make. It can produce a mistaken output, or it can make verbal errors (which are corrigible by feedback mechanisms) in the expression of the output. But it cannot misidentify the output of the analyzer - because it does not identify the output at all. What makes an output the output it is is what it goes on to produce in the speech centre, barring verbal (speech centre) errors. Thus the machine can be wrong about what is actually before its "eyes" (as in illusions, hallucinations, and other misidentifications) and it can make only verbal errors in uttering its reports of perception, but it cannot misidentify the output which comes from the analyzer, or in other words that which it seems to see.

Suppose that instead of making its reports in the form "I see a man approaching", it always wrote "I seem to see ... ", or "it is just as if I were seeing ... ". Reports in this form would disavow responsibility for fraudulent input or mistakes in the analyzer, and hence would be infallible barring only verbal errors that can be corrected.

This should not be taken to mean that the change in the form of words changes what is going on; the switch to the "I seem" idiom does not ensure that a particular thing is being done (e.g., a report is being made about output, rather than about input or the outside world), but that what is being done is to be interpreted in a certain way. Whatever the form of words, whatever the sequence of printed symbols, what is printed will be an expression of the output; the form of words is just being used as an indicator that one is to discount discrepancies between output and outside world. One could just as well leave reports in the "I see an X" form, and attach a small sign to the machine, saying "This machine is not responsible for fraudulent input or errors in input analysis". Thus it would be a mistaken extrapolation from this example to say that for human introspective reports to be immune from such errors, the reporter must intend his report to be an expression of some output, and that this intention would be fulfilled by his use of the idiom "I seem to see ... ". From this one could argue that infallibility depends on what Austinian <u>speech act</u> occurs, and that the idiom "I seem to see ... " does not ensure that the proper speech act (which would be intentionally expressing the output) occurs. The question of intention, however, cannot arise here; with whatever intention an utterance is made, it will be an expression of the output of the human analyzer (which is not just a visual analyzer), and if it is considered as such it is immune from error relative to the outside world. In fact, of course, when we intend our utterances to be immune in this way, when we intend that others judge them in this light, we frame our expressions in the "I seem to see ... " idiom. In using the "I seem to see ... " idiom, one is not intentionally expressing the output of the human analyzer - one does that, whatever the intention, and <u>as</u> expressions of output one's utterances are immune from these errors - rather, one is intending that one's utterance be <u>taken</u> as <u>only</u> an expression of output.[1]

One point that emerges from this is that there is no gain in infallibility if the speech centre computer is given a randomly established system for the production of typographical sequences instead of being programmed to write English. In either case irrefragable correctness of the printed messages is relative to the programmed rules of language, whether these are rules of English usage or of some randomly built "private" language. Translated as a partial answer to "private language" arguments, the point is this: the certainty of correct application of private

1 I am indebted to Dennis Stampe for pointing out these questions about speech acts and intentions.

words hinges on the user's being the final arbiter as to the correctness of a particular use relative to his private rules; whereas the correctness of reports in public language hinges on the user's being the final arbiter (feedback follower) as to the correctness of a particular use relative to the public rules. There is the same dependence on linguistic rules in both cases. So there is no gain in certainty in the programming of personal private terms but only a loss in objectivity and communicableness.

When this explanation of infallibility is projected onto the human mechanisms of perception, the problem of certainty is dissolved. There is a line that is crossed by the signals travelling from sense organs to tongue such that any uncorrected errors before the line change that which is to be expressed, and any uncorrected errors beyond that line are purely verbal mistakes.[1] In the envisaged machine the line can be drawn between analyzer and speech centre computer. Drawing the line in the human brain may prove to be a more difficult undertaking. But is this not a gap or turning point after all? Is this any better than the

---

1 Smart says that introspective reports differ from all others in having a "private logic". " ... if a person (a) knows the language and (b) is sincere, is not lying, then as a matter of logic he must have the immediate experience that he reports." (Op. cit., p. 99.) True, but how does Smart propose to explain this? Explanations come to an end, but not this soon.

stimulus checking mechanisms I have imputed to Smart's view? The crucial difference is that the stimulus checking mechanism is supposed to distinguish qualitative similarities and differences; its reports do not say what the qualities are, but just that there are these differences and similarities. The speech centre computer, on the other hand, has nothing to do with the ascertainment of qualities, but with the performance of programmed duties. The formation of a particular message by the speech centre is not triggered by ascertainment of any quality of the analyzer's output, any more than a lightbulb turns on because it has ascertained a particular quality in the incoming electricity. This is not merely a matter of ways of speaking. If one talks of qualitative differences and similarities, then one is saying that what makes a particular output the output it is are its physical qualities, and then the problem of ascertaining these qualities arises.[1] But what makes an

1 This is Ayer's view in The Problem of Knowledge: "To know that one is having whatever experience it may be, one must not only have it but also be able to identify it correctly, and there is no necessary transition from one to the other." (p. 68. See also p. 65.) If this is Smart's view, then his remark on the "private logic" of introspective reports is a contradiction. Ayer's view is wrong since whatever experience is identified (whatever output is correctly expressed in language) is the experience had. To give Smart his due, he rejects the idea that sensations have unanalyzable qualia by which we determine sensation location, but he never generalizes the point; he says only: "In particular, I see no need to suppose qualia as local signs." (Op. cit., p. 104.)

output the output it is, is how it functions in the computer, what is causes to follow it, regardless of its physical properties. To use the notion of content developed earlier, what is uttered is not a description (even a topic-neutral description) of a neural signal, since the signal as a message may *be* a description. The utterance made is the message uttered, the content expressed, and not at all a description of the neural vehicle of the message. He who reads this sentence aloud is not uttering a topic-neutral description of the marks on the page. When there are descriptions to be made there is the possibility of more than mere linguistic error; when messages are to be uttered there is no such possibility.[1]

It must surely seem that the time has come to posit the identity of thoughts or experiences with neural signals, in particular with the neural signals that are the output of the analyzing part of the brain, the input of the speech centre. For after all, what one would ordinarily call the content of the thought I have called the content of the signals. The descriptions of inhibition at the different levels

---

1 Again to give Smart his due, he does say that the brain processes which are conscious experiences are "those which can be *causal conditions* of our uttering ... introspective reports." (P. 103.) But it seems that Smart has not considered the implications of this remark *vis-a-vis* his notion of topic-neutral description.

of motor implementation in the speech centre were neatly parallel to introspective descriptions of talking and thinking to oneself. One ordinarily says that one can say what one's thoughts are with certainty - barring verbal slips or misuse of words - and I have shown how this infallibility works. But there are objections that can still be raised. I have said that one expresses the message of the output of the analyzing part of the brain when one utters a report, but this still leaves room for a dualistic argument.

One could say that the experience or thought is what is <u>reported</u> when the content or message of the output is <u>expressed</u>. For it is admitted that it is not the output that is reported - it is merely expressed - so the thing that is reported must be something else, the experience or thought. As an ontological argument, this is simply poor philosophy. One would be saying that since we do in fact say that we are reporting something, there must be something that is the thing we are reporting. But we also talk about the Average Taxpayer, so must there be something that is the Average Taxpayer? We need not talk of reporting things or of Average Taxpayers, so the fact that we do talk this way is no proof of the existence of the supposed objects. If I decide to talk about the things the perceiving machine reports, must there then be some things in addition to its physical parts? From the point of view of ordinary language, there is nothing wrong with saying that the thought is what is reported. In fact this is much more natural than saying the thought or experience is expressed. But one cannot make ontological capital from this observation. It is perfectly true that in the analysis of certainty offered, there is no thing that is the thing reported by the utterer, hence there is no thing that is well referred to by the ordinary word "thought", but that does not mean that the analysis has left something out. It is wrong to posit the identity of thoughts with neural signals or states, not because thoughts are something else, but because "thought", in the sense of "that which is reported", does not refer to anything. It would be wrong to say that the present King of France is de Gaulle, not because the present King of France is someone else, but because "the present King of France" does not refer to anything. And just as it would be wrong to say "but if 'the present King of France' does not refer to anything, then France is without a ruler", it would be wrong to say "but if 'thought' does not refer to anything, people do not think".

CHAPTER 8.

Awareness and Consciousness.

19. The Awareness Line.

The line beyond which only verbal mistakes can be made I shall call the awareness line. This concept requires considerable elaboration. The first step is an examination of the meaning and use of the ordinary word "aware", with a view to proposing a non-ordinary substitute. I am accepting as plain that there is some concept something like the ordinary concept of awareness that would be useful and illuminating in theoretical talk about human beings. I shall argue that the ordinary concept of awareness, as revealed in the tangled patterns of accepted usage, dubious usage, and misuse of the word "aware", does not have the restrictive rigour that one might suppose, or certainly that one would demand of any theoretical concept. In particular, I shall argue that the boundary between what we are aware of and what we are not aware of, certainly an essential feature of any sound concept of awareness, can be drawn only if awareness is attributed only to speaking creatures.

Consider a cat stalking a bird on the lawn. We ask, ordinarily, if the bird is aware of the cat, of the danger, only of the worm in the soil, only of the sound of the worm.

If the cat gets the bird we may decide to answer the first question No and perhaps also the second. But what about the third and fourth? Or if the bird escapes, was it aware of the cat, or just of the danger? We cannot ask the bird, which is the usual method of confirmation of awareness, but then is the question unanswerable, except by some arbitrary and misleading ruling?

Suppose I am tapping my fingers on the desk. I may be aware of this or not. One has to ask me to find out, even if I am staring at my fingers, since I may be "lost in thought". I drive a car for a hundred miles, altering course thousands of times, and at the same time chatting with a friend. Am I aware of all the kerbs, poles, corners and other obstacles in virtue of which (somehow) I altered course? Am I aware of them as kerbs, poles, etc.? Must I be aware of the little girl who darts out as a little girl before I can apply the brakes?

An insect gets through life with very simple neural circuitry that enables it to avoid obstacles, to mate, and to find food. Consider an insect that makes a successful detour around a tree; certainly it would be fanciful to suggest that the insect was aware of the tree as a tree; what do insects know about trees? But it would also be fanciful to suggest that the insect was aware of the tree as an obstacle, the way a human can be aware of something, in an abstract way, as an obstacle, a thing that frustrates one's plans, a thing that must be got around. One does not imagine the insect thinking "Ah, here comes another obstacle" or noticing, afterwards, that it had not been so much aware of the tree as a tree, but only as an obstacle.

One might say it is at least certain that if insects are aware of things they are not aware of things in a way at all like the way people are aware of things. One is then tempted to add: we cannot <u>know</u> what insects are aware of; if only the insect could tell us! But what of the little mechanical cars which have been made that learn how to avoid obstacles or traverse a maze? Is one tempted to say here: "if only the car could tell us!"?

"Aware" in ordinary use is at least quasi-intensional, to use the current idiom. We say that we are aware of things as certain things, under certain descriptions. There is no need even to lean on the "as" as I did above. The restrictions on substitution are made perfectly clear in an exchange like: "were you aware of that little girl?", "No, all I was aware of was a swift blur of motion". There are related uses for "see", "notice", "pay attention to": "All I saw was a blur"; "Fred, you haven't heard a word I've said!" - "Quite right, Ethel; all I heard was an incessant babble of noise";

"I noticed the sign, but I didn't pay attention to what it said". One might analyze these uses in terms of awareness: he saw the girl, but was aware of her only as a blur; he heard the words, but was not aware of them except as noises. But are there not three possibilities? Fred may be aware of what Ethel is saying (he may be listening and paying attention), or he may be aware only of a babble of sound (like listening to a babbling brook), or he may be entirely unaware of her noise-making (as of the clock ticking). He may not be aware of it as a meaningless babble at all; if he were, one might expect him to be able to say something like "listen to that babble!". If I am walking along the street lost in thought, and a friend passes by on the other side and smiles at me, I may be aware of my friend, or merely aware of some moving person or blur (which would be strange, unless I were tearing along), or simply entirely unaware of the situation under any description. The latter possibility is not at all the same as being convinced or under the impression that I am alone on the street. If I am looking for a button, and it is right before my eyes, if I were aware of it only as a shapeless <u>it</u> or unformed blur, I would certainly focus on it, take a second glance, just in case it was the button. If I do not do this, I must be totally unaware of it.

One possible definition of awareness is that one is aware of whatever information contributes to the control of behaviour. The strict way of determining this would involve ascribing content to neural signals, but until such time as this is possible, a close intuitive approximation could be made by considering the behaviour in question and attributing awareness of the minimum information that would suffice to control the behaviour. Thus the automobile driver would be said to be aware of <u>things</u> (which in fact are signposts, kerbs, parked cars, etc.) and this would suffice for him to avoid these things. This would fit some of the uses of "aware" quite well, but there are difficulties. The insect is then aware of the thing (tree) that it flies around, and the bird is aware of the edible, noise-making thing (the worm). People do say that animals are aware of things, but should they? Worse still, on this definition missile guiding computers are aware of the trajectories of their missiles, and automated orange sorters are aware of the size and softness of the things (oranges) that are sorted. It would be unenlightening to accept the definition and then, by edict, restrict it to animals or just to human beings. That would be like giving a definition of genius in terms of the results of some I.Q. test and then further restricting it to a list of one's personal friends and heroes. And even in the case of human awareness this definition runs into difficulty. Should one say that the casual walker is aware of the proprioceptive information from the soles of his feet and his joints and inner ear without which he could not walk? Consider a man who reasoned thus: I must have been aware that my glass had reached my lips, or I wouldn't have tipped it. One would, on this definition, be aware of things that one could not report, except by inferring that one was aware of them from the behavioural (and ideally, neurophysiological) data. It has been shown that table tennis players rely on the sound of the ball striking the table even more than on the sight of the ball. Imagine a table tennis player saying "I had no idea I was aware of the sound - except as a meaningless din - but now I see that I must have been, all along." I suggest that the person who said this would immediately <u>start</u> being aware of the sound as more than a meaningless din, and that his game would suffer, just as the typist or pianist who pays attention to his finger motions comes to a grinding halt.

Another possibility would be to restrict awareness to the "highest level" or "highest value" information in some way. Then one is aware that one is walking, but not of all the subordinate information; one is aware of where and when the ball hits the table, but not of the jots of auditory and visual information that go to make up the higher information. But this does not rule out the orange-sorter or the insect, and furthermore can I not be aware of the pressure on my soles when I walk if I want to?

The notion of awareness that is being sought clearly has it that the phenomenon is detached or detachable, with certain restraints, from behaviour control. The notion of information that controls motions is useful, particularly if it can be applied equally to animal behaviour, certain human behaviour, and the operation of sophisticated information-dependent machines, but this is not the notion of awareness that is "intuited" when one considers the "introspective fact" that one knows automatically or without inference what one is aware of.

A definition of "aware" that fits this aspect of our ordinary talk of awareness is that one is aware of that of which, at any moment, one can make an infallible report, barring verbal mistakes. But this formulation cannot stand, since, in its reference to <u>reports of</u> ... , it raises the problem of infallibility laid in the last chapter. It should be emended to: one is aware at any moment of that to which one can give expression at that moment, barring verbal mistakes.[1] Then, what one is aware of are the contents of those signals that, at any moment, cross the line beyond which only verbal mistakes are possible. Hence, the awareness line.

20. The Status of the Awareness Line.

"Aware" defined in this way is patently not the same word, in all its trappings, as its ordinary language model; "are you aware that Jones just died?" and "Fido, aware that his dinner was ready, ... " are contexts into which the newly defined word will not fit. Furthermore, many philosophers would want to say that we are aware of our thoughts, and as I argued (in §18) it is best not to say that thoughts are the messages carried by the signals crossing the awareness line. If, however, the definition did fit all the peculiarities of ordinary usage, the term defined would still exhibit all the puzzles. There would be no gain. The definition fits one aspect - an important and separable aspect - of our ordinary talk of awareness, and that is enough. I shall indicate the non-ordinariness of the term with a subscript: "$aware_1$". And since the notion of information that controls motion was seen to be useful, although confusedly

1 It would be simply redundant to say "incorrigible or infallible expression ... barring verbal mistakes".

mingled with the notion of $awareness_1$, I shall say that animals - and some machines - are $aware_2$ of the controlling information.[1] The insect is $aware_2$ of the (information that there is an) obstacle, and the bird is $aware_2$ of the (information that there is a) worm.[2] Once this distinction has been made, it becomes clear what one could mean by saying that animals are not aware of things in at all the same way humans are; animals are only $aware_2$ of things, which is saying very little, since machines can also be $aware_2$ of things. The temptation lapses to say we cannot know how animals are aware of things, and if only they could tell us. If animals could tell us, they would be $aware_1$ of things, which is entirely different. The basis of $awareness_2$ is behaviour control; the basis of $awareness_1$ is verbal expression. The concepts are totally different, and it is when the halo of intuitions around one of these merges with the halo of intuitions around the other, that confusion results. One can say, using this distinction, that insects are simply not $aware_1$ at all, and then the question cannot

1 Jonathan Bennett uses a similar stipulative definition to discuss what bees can "know", in Rationality, 1964, p. 18. (See ahead, §35.)

2 I am ignoring for the moment the practical difficulties with quasi-intensional ascription encountered with Doggish in Chapter 6. The bird is not $aware_2$ of the organism of genus *Lumbricus*; it is aware of the "worm" (in scare quotes).

be asked: But if the insect was not aware of the tree, how did it know enough to fly around it? The insect was $aware_2$ of the tree.

I do not want to suggest that the concept of $awareness_2$ is tailor-made for the rigorous use of neurologists, ethologists or other scientists. They may devise much more useful and rigorous concepts. The chief value of "$aware_2$" is in putting off these who would insist that awareness is a prerequisite of regular appropriate behaviour. The term is at least harmless, which is a step in the right direction.

The notion of $awareness_1$ then is restricted to creatures that can express, or in other words, speaking creatures. This is not just because humans have traditionally been considered to be the only creatures with awareness, but because this notion of awareness is defined in terms of verbal capacities. Non-speakers could no more be $aware_1$ than they could be guilty of mispronunciation. Of course any machine that had a speech centre computer attached would be $aware_1$ of the input of its speech centre, which is the output of the rest of the machine. This may seem to be an intolerable situation, but only if one clings to the folklore that has accrued to the ordinary word "aware". There is certainly nothing wrong with a machine being $aware_1$ of its output, if <u>all</u> this means is that it can express its output correctly barring verbal mistakes. But if that is all the word means, there is still a great deal to explain or explain away, for I wish to show that there is no important residue in the ordinary concept of awareness that is not subsumed under either awareness$_1$ or awareness$_2$. There is no room, I wish to show, for a concept of awareness$_3$, which would apply only to humans and rule out machines.

I have said that we are aware$_1$ of the contents of those signals that cross the awareness line, but I do not claim to have discovered some special seat of consciousness, some organ of the brain set aside by nature or God. The awareness line is not a gap in the gray matter or a membrane, and it does not enclose, after all is said and done, the "private theatre" of the mind. It is a projected line that signals can be said to cross. This projection is dependent on the concept of awareness$_1$, the concept of what, at any moment, a speaking creature can say, allowing for verbal errors. There are speaking creatures, and they can be observed to correct themselves occasionally when they make slips of the tongue or other verbal mistakes; if one decides to talk about the standard by which a speaking creature corrects his verbal mistakes, then one needs the concept of awareness$_1$. The decision to talk about this standard was made long ago, when the concept of awareness developed, and now the problem remains to separate out the right concepts from the wealth of theory embedded in the ordinary concept, or our "intuitive" concept.

The awareness line is thus in one sense not a physical feature of the brain, any more than the Equator is a physical feature of the earth. The Equator is a line projected on the earth considered as a spinning spheroid with a more or less stable axis. Once men had the concept of a great circle equidistant from the poles, they could determine its location, which is not physically arbitrary. Once we have the concept of a line beyond which only verbal mistakes can be made, beyond which any uncorrected neural malfunction or misprogramming will result in a verbal mistake, we can determine its location in the brain given the necessary auxiliary tools and knowledge of the brain. But such a discovery, like the discovery of the Equator and unlike the discovery of North America, is dependent on the concept.

The determination of the location of the awareness line might prove to be a particularly sticky problem. There is no particular reason to suppose that the neurophysiologists, in their separation of a part of the brain as the speech centre, have already drawn the line. For there may be levels of linguistic programming outside the boundary of the traditional speech centre, and this would locate the awareness line outside it. There is also no a priori reason for the awareness line to be regular, stable, and unbroken; it may prove to be better to consider it as a scattering of convoluted, mobile line segments. One way of determining it would depend on the ability to identify feedback loops. At the end of the deepest level of vocal feedback loops is the awareness line.

This means of course that until the awareness line is actually located, my remarks about $awareness_1$ can be construed as circular. To say that one is $aware_1$ of a certain signal content once that signal has crossed the awareness line is, on my definition of the awareness line, tautological. But before the Equator was carefully mapped, so was the statement: one is in the Southern Hemisphere once one has crossed to the south side of the Equator. This circularity does not detract from the value of the concept; it is merely a manifestation of the overall dependence of physicalism on scientific corroboration. This dependence is the strength of physicalism. Unlike traditional non-physicalistic theories of mind, which can be neither corroborated nor disproved by scientific observation - no scientist is ever going to come across a transcendental Ego, a noema, a sensum or an idea - a physicalistic theory of mind has the opportunity of becoming accepted into the realm of science.

It is clear that the notion of awareness$_1$ has been laundered of three different sets of connotations that ordinarily accompany the notion of awareness. First, I have allowed for no pictorial or imagistic connotations; there has been no talk of colours, images, appearances, and so forth. In applying "aware$_1$" to language-using machines I have not suggested that these machines have a rich inner life of psychic imagery, even though the machine described was a perceiving machine. Since this notion of inner imagery is tenacious, it will be dealt with at length in Chapter 9. Second, I have not left any room for the sort of creative marshalling of thoughts that is generally supposed to go on "in consciousness". Signals simply arrive at the awareness line, and no mechanism has been suggested that might arrange these, infer from these, consider these, or jump to conclusions on the basis of these. All one can do with these, to put it crudely, is say them. The notion of thinking as an active creative process will be treated in Chapter 10. Third, I have separated awareness$_1$ from control of behaviour, and this matter will be considered in more detail now.

21. Awareness and Control.

In a human being, if a signal does not cross the awareness line, he cannot express its content; he is not $aware_1$ of its content. But that does not at all mean that such a signal may not go on to trigger a very useful circuit in the nervous system. A simple reflex like blinking or pulling one's finger away from something hot is activated by circuits that never come close to the speech centre. Signals are sent on to the higher parts of the brain, and a signal to the effect that one's finger is touching something hot can thus eventually cross the awareness line, but by that time the finger has already been jerked away; coming into $awareness_1$ has not contributed to the behaviour. One can become $aware_1$ that one is blinking, but one almost never is. In between simple reflexes and activities of which we are quite definitely $aware_1$ there are activities of great complexity whose neural circuits never become involved with the awareness line. (By activities of which we are $aware_1$ I mean, of course, activities about which there are signal contents of which we are $aware_1$. For the sake of smoothness I shall now partially revert to ordinary usage in allowing $awareness_1$ to be of things, activities, events, etc., since in each case the phrase can be construed as $awareness_1$ of the content or message of signals to the effect that such things, events, activities are as they are - or seem to be.)

An accomplished pianist can play Chopin beautifully, "with his mind on something else", and need not be $aware_1$ of the notes on the page, the sounds of his playing, or the motions of his fingers. He must, of course, be $aware_2$ of these. Is there anything strange about this? For could one not build a machine that read music and then played it (a somewhat sophisticated player-piano)? There is no temptation to suppose that it is anything more than $aware_2$ of what is going on. The man who drives for miles "all the time immersed in thoughts of his home" neatly steers clear of all obstacles, changes gear, brakes, accelerates, and he need not be $aware_1$ of any of this. If someone asks him what landmarks, buildings, cars, etc., he has just passed he will be unable to say, and not necessarily because he was $aware_1$ of these things but already has forgotten, but because he was only $aware_2$ of these things.[1] Of course if one asks him what at any moment he is passing, he can immediately become $aware_1$ of what is passing, and give reports in high detail. That is, the question acts as a

1 This suggests that one can only remember things that one was, at the time, $aware_1$ of. But there is more to be said about memory than that. Memory in its various guises will be referred to in passing in the chapters that follow.

stimulus for the behaviour of reporting visual experience, which without the question would not be particularly appropriate, and might be inhibited to allow for the more pleasant behaviour of thinking about home to occur. One cannot think about two things at once, but one's brain can control two or more activities at once. We say we do many things without thinking, but we do not do these things without neural activity to control them.

The output of the human brain, unlike that of the perceiving machine, can come from various sources, so the input into the speech centre may not always be the output of the visual-analyzer part of the brain. There is no immediately apparent reason why the input could not switch back and forth between visual signals and, say, memory signals, at a great rate of speed, but in fact this seldom if ever happens. One has fairly long, uninterrupted sequences of one kind of signal. Shifts in the output occur whenever appropriate. It would be rare for a man to drive long distances without occasionally being $aware_1$ of his driving, and similarly the pianist would not remain $unaware_1$ of the notes, the sounds, or his finger motions for long. In particular, if he made a mistake, the negative feedback would no doubt shift him to $awareness_1$ of what he was doing.

This suggests that awareness$_1$ does have some efficacy in behaviour control. What is the value of paying attention if not to control one's behaviour better? Surely when one pays attention to something, one is aware$_1$ of it. But it is also clear that simply bringing a signal across the awareness line, "into speaking position" so to speak, could have no beneficial effect on behaviour control. So there can be no strong relation between being aware$_1$ of something, and improving one's control of related behaviour. But there could be a contingent and coincidental relation.

There is no doubt that we bring activities into awareness$_1$ to correct them or improve them. If I keep fumbling a trill on the piano, I start paying attention to the particular motions of my fingers when trilling. When learning to drive, one is very much aware$_1$ of raising the clutch, shifting gears, looking in the mirror, and so forth, although these activities will eventually become "automatic". It is also clear that we are inevitably aware$_1$ of the sights, noises, and other sensations (to use the word unphilosophically for the time being) that are particularly bright, sudden, loud, acute, bizarre, unexpected, or otherwise outstanding. It would be appropriate for us to be aware$_1$ of these sensations if our awareness contributed to better coping with these sensations, since the outstanding sensations are usually the ones that make the most difference to the well-being of a person. I suggest that becoming aware[1] of these sensations, or of the activities that one must pay attention to, is a contingent by-product of a shift in controls that occurs in the part of the brain on the far side of the awareness line.

The signals that cross the awareness line are high level (if not the highest level) signals of the brain. There are no direct (single axon) connections between retinal or motor neurones and the speech centre. So one would expect that whatever neural activity was monopolising the highest reaches of the brain would be the activity whose constituent signals - or signals triggered by the constituent signals - would cross the awareness line, if any signals did. When one considers the brain as a controller of non-verbal behaviour - as it often is - one can see that any input into the speech centre would be a by-product of the controlling activity. If controls of behaviour could be handled by lower level organizations of the brain, then the higher reaches would be free to deal with other matters. But if the higher reaches must be used to control the behaviour, then the signals contributing to control at the higher level would be the signals that would cross the awareness line.

When one is first learning how to perform some difficult operation, like tying one's shoelaces, the activity must be controlled by the highest levels of the brain. The whole range of afferent input must be analyzed and the crucial features picked out. Verbal and ostensive instructions from others must be coordinated with the visual input; stored information must also be brought into play with the panoply of externally stimulated signals. Once the bare essentials (mainly proprioceptive signal patterns from fingers) have been culled out, and an order of motor operations established, the whole organization can be packed off into specialized controls, which require feedback from only the fingers; one can tie one's laces without even looking. Eventually one can tie one's shoelaces "without even thinking". At this stage the controls are like an automatic pilot, and do not involve the use of the higher levels of the brain.[1] Now if the automatic pilot failed to work properly (because it had been faultily programmed in the first place, or because the laces were too stiff, or slippery or short) the failure or the organisation would shift the problem back up to the higher level again.[2]

---

1 My use of "higher" is subject to revision. The cerebellum (is it "high" or "low"?) is perhaps the locus of "automatic pilot"-type controls. See Ruch, "Motor Systems" in Stevens (ed.), Handbook of Experimental Psychology, 1951, ch. 5.

2 The hypothesis is that there would be negative feedback from the control organization, indicating that it had

This point could be generalized in the ordinary observation: whenever we have difficulty in doing something, we concentrate on it. I suggest that concentrating or paying attention to what we are doing is first a matter of bringing the control of the activity into the higher levels of the brain, where there is access to the full range of stored information and incoming afferents, and only incidentally or contingently a matter of bringing the problem into awareness$_1$. It so happens, I am saying, that whatever the brain is concentrating on, we are aware$_1$ of, but that concentration in the sense of bringing the highest level of the brain to bear on a problem, is separable from awareness$_1$. There is no reason why animals cannot concentrate on what they are doing, or pay attention (through facilitation) to particular parts of their afferent input, but that does not mean they are aware$_1$ of anything; they are aware$_2$ of many things, and of these they may be facilitating some, but they are aware$_1$ of none.

Another facet of this coincidental relationship between awareness$_1$ and control is the fact that if we are unaware$_1$ of things, they cannot bother us, i.e., prevent us from using our brains for other matters. A sore foot may be so sore

---

not completed its cycle, and this would shift the problem. For a detailed analysis of this kind of function, see Galanter, Miller and Pribram, <u>op. cit.</u>

that I cannot "get my mind off it", and am almost continuously $aware_1$ of the pain. But if I am able to concentrate on other things, I will not notice the sore foot at all. If someone then asks me if my foot is still sore, I may become $aware_1$ of the pain, and answer Yes, but this does not show that I have been $aware_1$ of the pain all along, nor that the pain has disrupted or prevented other high level neural activity. A comfortable chair is not one that is constantly impressing one with the pleasure of sitting in it, but one that is not noticed at all, that allows one to concentrate on other things.

To say that one is not $aware_1$ of something is not to say that he is $aware_1$ that something is not. When I am lost in thought and $unaware_1$ of the things in my visual field, it is not like being blindfolded; everything does not go black or blank. If I am $unaware_1$ of the chair I am sitting in, that does not mean that I am $aware_1$ that I am hanging in empty space. Or, if someone asks me what colour the walls of my friend's study are, I may say that I have never noticed; and it may be perfectly true to say that if they had been bright red I would have noticed, but that does not entail that I noticed they were beige, nor that I was $aware_1$ that they were colourless.

It might be objected that it would be much better, much more in line with everyday ideas, to identify those things that one is aware of as those things that are foremost in one's mind, since whenever I have spoken of someone being unaware$_1$ of something (like the notes of music), it has only been because something else was foremost in the person's mind. But I wish to hold that it is only a contingent happenstance that what we are aware (not aware$_1$) of somehow has temporary hegemony over the rest of our "mental" activity, and hence I wish to define non-ordinary "aware$_1$" so that this is explicit. There is a choice: if what is essential about awareness is held to be its relation to behaviour control, then it is only contingent that one can "say what one is aware of", that one can express the content of which one is aware$_1$; if what is essential is held to be that infallibility of expression, then it is only a coincidence that we are aware of the information of foremost importance to the control of present behaviour. On the former view, animals are once again brought into the fold of creatures with awareness. This would be all right, provided that no one then was tempted to say "if only the cat could tell us!" Is there room for still a third type of awareness? Awareness$_3$ could be defined as being of the highest level controlling signals. Then one might be justified in saying of a cat engrossed in playing with a feather, that it was only $aware_2$ of its position on the rug, the sunlight, the position of its tail, but $aware_3$ of the floating feather. But this, I suggest, is simply carrying the game too far. The essence of awareness is $awareness_1$, and the countenancing of other uses is merely paying lip service to vestigial preconceptions.

Consider a juggler, balancing himself on a large rubber ball and at the same time balancing a spinning ball on his fingertip. Does his behaviour indicate which of these quite independent activities he is aware of (the ordinary word)? Our everyday experience would suggest that one of these activities, whichever is more difficult for him, must "occupy most of his attention". He may be concentrating on the spinning ball, while he maintains his balance automatically. But is it not quite possible that he is such a good juggler that "his mind is elsewhere", that he is thinking of the blonde in the second row and is quite unaware of either of his juggling activities? Of course if anything started to go seriously wrong, he would immediately become aware of whatever it was and proceed to correct it, but he could correct minor errors automatically.

Now consider a seal performing the same trick. Is there any reason to suppose that the seal is aware of either of its activities? It could not say "I am quite aware that

I am earning my fish by balancing on this infernal ball and spinning another on my nose". I grant that the seal may not have relegated one of these activities to sheer automaticity, and hence the higher parts of its brain may be involved in controlling it, but should one then say that the seal is aware of that part of his act? Would it make sense to suggest that the seal's mind was elsewhere? Through conditioning, the seal may well be in a "state of expectation" for a fish, in that its neural activity has stimulated salivation and certain afferents associated with receiving a fish would terminate some neural state, but since the seal has no means at all of expressing the content of the signals that occur, no sense can be made of the suggestion that the seal could intuit or be aware of what it was thinking. The little man - or in this case, the little seal - who was supposed to sit watching the screens of the senses, introspecting the thoughts and so forth, has been replaced by the programmed verbal expression of messages carried by neural signals. Lacking a programmed language centre, the seal cannot be $aware_1$, cannot introspect, cannot intuit. But this does not mean that the seal is doomed to live an unhappy life of uncomprehending darkness, always wondering what is going on and what it is doing. It simply means that such talk is inapplicable to seals; seals get from here to there, eat, have young, lead normal seal lives, without the help of language, and thus without the trappings of awareness. Only a creature that can be aware of something can be sadly unaware of something. People struck blind are depressed by the loss; blindness does not bother stones.

It is easy to confuse awareness in its sense of awareness$_1$ with the notion of behaviour control, and to extrapolate from the fact that one must be aware of something in order to say it (express it, report it) to the untruth that one must be aware of something in order to do anything with it. Thus Kenneth Sayre says that "we would not say under any ordinary circumstances that we recognize an apple, or some other object, but were aware of no such object".[1] Of course we could not say we recognized an apple and yet were not aware of an apple, since in order to say anything, we must be aware of it, but what of machines or animals that cannot say anything? They can fulfill all the functions of recognition short of saying they recognize; does this bar them from recognizing? It is not that one must be aware$_1$ of something in order to recognize it, but that one must be aware$_1$ of something in order to say that one recognizes it. The success of "subliminal" advertising, of which we are

1 "Human and Mechanical Recognition" in Sayre and Crosson (eds.), The Modeling of Mind, 1963, pp. 157-70.

unaware (on even the intuitive interpretation of awareness) is strong evidence that awareness and higher behaviour control are only contingently connected.

22. Awareness and Speaking.

An objection that can be raised to this account of awareness is that we often say "He isn't aware of what he is saying". Since I have said that we are $aware_1$ of what we can say, this would be a counter-example from ordinary discourse that must at least be considered. The idiom in question can mean two different things: (1) he has not thought through the consequences of his remark; he does not see the undesirable import of what he is saying considering the present situation; or (2) he is reciting automatically or babbling. (1) is no counter-example since I have not said that one must be aware of or must realize or must have figured out all the connotations and consequences of one's remarks. One can be perfectly aware of what one is saying in the narrow sense without seeing the dire effect that will follow. (2) necessitates some further refinement of the notion of $awareness_1$.

The hypothetical perceiving machine that served as the model for my account of certainty was a single-purpose machine. Hence only one pathway was delineated for the flow of information leading to the printed messages. But the human brain is a multi-purpose mechanism; in addition to making reports, it produces questions, lies, orders, recitations, and formulae required by etiquette. The problem raised by (2) can be dealt with if the perceiving machine is given the additional function of reproducing written statements - of reciting, in other words. This could be done by placing the material to be reproduced on the printed tape in front of the television cameras, or by using a separate specialized scanner. In either case the output from the scanner would not need to be fed through the entire speech centre computer; that preliminary part that composed sentences as expressions of the input data could be bypassed and the data could be fed directly to that part of the speech centre computer that controlled the printing of letters and spaces on the output tape. There would still be feedback for correcting printing errors, but it would not carry beyond that to the correction of sentence structure, word choice, and so forth. If such a direct line were established, the bypassed part of the computer, which could be called the sentence-synthesizing or synthesizing part, would be free to carry on other activities, provided its output was blocked or inhibited before reaching the printing-control part of the computer.

I wish to apply this account to the example of a child learning to read. At first, reading aloud is far from automatic. The child looks at each word, determines what the word is, and reports it. I suggest that there is a subtle difference between reporting what words are on the page, and reading them off, the difference being a matter of how automatic the activity is. The child's first successes I would want to describe as parallel to the perceiving machine's reports to the effect "there is (seems to be) a card, and on it are marks that spell out 'cat'". The development in the child of the ability to read effortlessly suggests the development of a more direct path between visual stimuli and utterance, the development of a system of recitation similar to the one described for the perceiving machine plus special scanner. The independence of such a system would depend on the difficulty of the material to be recited, and the firmness and correctness of the programming of the system.

Once such an automatic pilot had been firmly established, it would even be possible for the synthesizing part of the speech centre in the child to carry on its normal activities, provided they were inhibited before the point at which they triggered the formation of utterances. A consideration of the relevant observations about learning to read supports this type of story. A child who lets his mind wander when reading aloud is soon unable to go on. A new or difficult word brings him to a halt. Adults often find that they have read a page without taking in a single word or idea; their eyes have followed the words and sentences, and their lips may even have mouthed the words, but nothing has sunk in; their minds have been wandering. This infuriating failure is unknown to the child; his reading activity is not yet so well-programmed that he can do this. This phenomenon is not to be confused with reading without comprehension, which will be mentioned later; this is reading without even awareness.

The adult can recite automatically, provided the material is not too difficult, or in a foreign language, and while he is doing this he may be aware of the expressions on the faces in the audience, or his wish to be through with the reciting, or whatever. Or he can inhibit the last rank neurones that stimulate the contraction of muscles and recite to himself, all the time being aware of something else. He need not have something written to recite. He may say, over and over again, some simple utterance like "1-2-3-4-5-6-7-8-9-10", or "I'm talking to myself, I'm talking to myself, ... ", and at the same time be aware of other things, especially, that he is doing just this, that it can be done! He cannot, however, say involved or non-repetitive things to himself, and at the same time be thinking of something else, unless he is reading something automatically. The child, I would venture to say (although it is hard to see how this could be verified now), cannot do this. Even the number sequence is not usually so automatic for him that he can say it to himself and be thinking of something else at the same time. I would say that it is physically, but only temporarily, impossible for the young child to be unaware of what he is saying out loud or to himself, but that the ability to do this comes naturally with the increase in the ability to talk, and especially to read.

Reading with comprehension is just a sophistication of hearing with comprehension. Given the general functional picture outlined above, the steps in learning to read with comprehension sort themselves out nicely. The programming of the speech centre - not necessarily for the perceiving machine, but certainly for human beings - involves two-way traffic of utterances; not only must the speech centre be programmed to synthesize utterances from its input from the rest of the brain, but to analyze incoming utterances - heard utterances - into signals of the same functional sort as the output of the analyzing part of the brain. Input and output must be compatible.

One comprehends heard sentences before one comprehends written sentences. Thus it is that the child must first read aloud, and, in effect, listen to what he is saying, if he is to understand it. This "listening" could of course be accomplished by proprioceptive information as well as by auditory information, provided the child was not deaf to begin with. The ability to understand what one is reading without even saying the words to oneself is again dependent on the establishment of short-circuit methods of analyzing the input into signals of the same level and functional type as the signals that cross the awareness line into the synthesizing part of the speech centre. Such an analysis is at least a necessary condition for understanding. (See §34.)

Now, if this analysis of being unaware (the ordinary word) of what one is saying is correct, should I say that we are or are not $aware_1$ of what we are saying when we recite automatically? First, it is perfectly clear that through <u>listening</u> to what we are saying automatically, we can be $aware_1$ of what we are saying in just the same way we can be $aware_1$ of what someone else is saying - and either with comprehension or as a meaningless babble. But we need not listen; we need only be $aware_2$ of what we are saying in order for it to be properly controlled and regulated. If one decides that it is essential that contents of which we are $aware_1$ must arrive at the synthesizing part of the speech centre, and not bypass it, then it would follow that we are not always $aware_1$ of what we are saying. Earlier I suggested that it might be best to consider the awareness line as a broken line; in that case, direct lines to the phonemic organizers, etc., could be said to pass through the gaps. If the question seems moot on the points of theory I have proposed, it also seems moot on "introspective" considerations. When I say to myself "cogito, ergo sum; cogito ergo sum; ... " over and over, at the same time thinking of Descartes and his oven, is it obvious one way or the other that I am or am not aware (the ordinary word) of both of these events? I see no particular reason here for not saying that I am $aware_1$ of two things at once, provided it is made perfectly clear that one of the things is a repetitive or automatic babble. In general, however, I think it would be clearer to reserve $awareness_1$ for those contents that arrive at the synthesizing part of the speech centre. The point is simply that once the situation is clearly understood, there is no strong reason to force the theoretical word "$aware_1$" one way or the other.

Another difficult question that can be asked about $awareness_1$ is: What am I $aware_1$ of when I utter a lie? a question?

an order? It is perfectly clear that I am $aware_1$ of the content of what I am saying, but that in itself sounds odd. We do not usually say things like "I was aware that 'How old are you?'" or "I was aware that 'Please pass the butter'". And to say that I was aware that ... , followed by the content of my lie, would seem to suggest that I thought the lie was the truth. But this is tying content too closely to the actual uttered words. Earlier I pointed out that a particular signal crossing the awareness line might well be ascribed the content: "apologize!", and that the speech centre followed through with "pardon me" or "please excuse me" and so forth. Similarly, the content of the signal crossing the awareness line in the case of a question would be much more like "get him to tell you his age (by asking him)", and in the case of a lie, "dupe him into believing that p (by saying that p)". The content, it must be remembered, depends on the function of the signal, and that function will be successfully performed provided certain appropriate changes occur in the environment. Except in the cases where the change sought is simply a matter of the occurrence of a particular phonemic string of sounds (as in reciting), the content is not simply the sentence uttered.

There is more to be said on the difference between lying and telling the truth, asking and reporting, supposing and stating, but I am postponing this for later chapters. One point that can be made now is that one need not be aware$_1$ that one is telling the truth <u>in order to</u> tell the truth; to suppose this would be to commit the fallacy scouted by Putnam (see §18); the computer need not discern the state it is in to be in that state. But there is more to it than that.

Another objection that might be raised is that often we <u>are</u> uncertain how to describe what we are aware of, and I have left no room for this in the definition of "aware$_1$". This uncertainty can be of two kinds. If I am presented with two drawn lines and asked which is the longer line, I may say I am uncertain. In such cases it is the input into the speech centre that is indeterminate with respect to which is longer; the difference in length is below the threshold of discrimination of my eyes, or the drawing is an optical illusion that systematically misleads the analyzing mechanisms. To put the matter in ordinary terms, it is not that the experience is clear and determinate and I cannot tell in examining the experience which is the longer line, but that the experience itself is unclear or indeterminate in this respect; the "report" does not say which is longer. (See Chapter 9.) The other sort of uncertainty is verbal, and owes its existence to the vagueness of the words programmed into the speech centre. "I'm not sure if

I am still in pain" can only mean "I'm not sure if the word 'pain' is not too strong for this case". No doubt the function of the speech centre mechanisms when stymied in this way by conflicting or vague programming is to produce sentences containing such idioms as "I am not sure ... " and "It's hard to say ... ".

Before turning to an analysis of the concept of consciousness, in the respects that it differs from the concept of awareness, I should perhaps stress that I have not attempted to provide a behavioural criterion for determining what people are aware of. One cannot infer from the fact that a person says "I see a rabbit in the bush" that there is a rabbit in the bush or that the person even <u>seems</u> to see a rabbit in the bush; he may be lying or reciting a poem, or something like that. All one can say is that <u>if</u> a person is making a report of what he sees or seems to see (and the perceiving machine could not do otherwise), then barring uncorrected verbal errors, his report is foolproof. We can in fact be informed as to what another is $aware_1$ of by his sincere reports, but at present there is no sure way of determining that his reports are sincere. One can determine that either a person is $aware_1$ that he seems to see a rabbit, or he is $aware_1$ that he is lying to that effect, or he is reciting, or he is misusing words in a way he could correct given a proper knowledge of the language. And it is obvious that we know this much.

23. Consciousness.

"Conscious" is not really an ordinary word at all, although it has almost become one. It is a theoretical word that has acquired all the currency, vagueness and ambiguity of an ordinary word. First, as Scriven for one points out,[1] "conscious" can mean something like (1) "awake" or "aware", or (2) it can be used to contrast animals that have this capacity for consciousness with lifeless objects that do not have it. It is not really paradoxical to say that only conscious animals can be unconscious. Stones are not conscious in sense (2), and hence can be neither conscious nor unconscious in sense (1). Or, a stone is unconscious in a different way than a man in a coma is unconscious. The simple way out of this ambiguity is to reserve the term "non-conscious" for stones and trees, and perhaps lower forms of animal life. This still leaves the problem of where to draw the line between conscious and non-conscious forms of life, and this can only be resolved by solving another problem: What is it to say that a being is conscious

---

1 "The Mechanical Concept of Mind", Mind, 1953, reprinted in Anderson (ed.), Minds and Machines, 1964, p. 33.

in sense (1)? (I shall use "$conscious_1$" for this sense and "$conscious_2$" for the opposite of "non-conscious".)

There is simply no consensus of usage here. Is a person $unconscious_1$ when he is asleep, or only when he is in a coma? There is a big difference, since when one is in a coma, even reflex action disappears, and only regulatory nervous activity continues. If being in a coma (or being dead) are the only ways one can be $unconscious_1$ then any animal with a properly functioning nervous system is $conscious_1$, and hence any animal with a nervous system, functioning or not, is a $conscious_2$ form of life. If one is said to be $conscious_1$ if one is awake, and not otherwise, some sticky questions may be asked about simple animals that do not seem to sleep, but aside from this, dogs and horses and other animals that sleep can be called $conscious_2$. This, I propose, is the best use of the word. A being is $conscious_1$ if its nervous system is not in the resting state (where afferents are inhibited, only reflex actions occur, regular brain waves are detectable, etc.). This leaves several problems. First, there are dreaming and sleepwalking and activity under hypnosis.

Little is known about the mechanisms involved in sleepwalking and action under hypnosis, so my remarks on these will have to be sketchy - although perhaps no sketchier than my remarks on neural functioning in general. Hypnosis seems to have the effect of shutting down or counteracting specific parts of the brain, and replacing their normal outputs (sometimes) with similar, but non-veridical outputs. Whatever else may be true of hypnosis, its effect is selective. In sleepwalking the effect might be much the same: parts of the brain are active while others are not. If this were so, then sleepwalking and action under hypnosis, and dreaming, which is probably no different from sleepwalking except in the inhibition of (most) muscular action, would simply be cases of <u>partial consciousness</u>$_1$. There is nothing in the proposed definition that suggests that the term must be applied in an all-or-nothing manner.

This proposed use of the term is fairly close to the accepted usage of neurologists. Geoffrey Jefferson, the neuro-surgeon, for example, would grant consciousness to animals other than human beings, and simply grade brains according to the complexity of operation.[1] W. Ritchie Russell says that consciousness is acquired gradually by the child, as its brain becomes programmed, and hence truly active.[2] But misgivings from two sources have militated against more than a wary acceptance of this notion.

---

1 "The Mind of Mechanical Man", <u>Brit. Med. Journal</u>, 1949, June 25, p. 1105.
2 <u>Brain - Memory - Learning</u>, 1959, p. 120 ff.

One source is the field of psychoanalysis, with its concept of the Unconscious. If the Unconscious is seen (at least metaphorically) as a functional area, then must not consciousness also be an area? Granit succumbs to these pressures when he says that there is no centre of consciousness in the brain; consciousness, for Granit, is "no more than a fringe on the pattern" at the highest levels of neural organization.[1] But if $consciousness_1$ is simply the physical state of being awake, then what can this talk of fringes mean?

The other, related, source of misgivings is our everyday use of "conscious" in the idiom "conscious _of_ ... ". Here the word poaches on the territory of "aware", and the ideas behind both words are mixed into an unmanageable _mélange_.[2] They can be neatly separated in the following way.

One is $conscious_1$ when one's brain is awake; one is completely $conscious_1$ when all of one's brain is awake; one is completely $unconscious_1$ when the brain is completely dormant (when one is in a deep coma) or when one is dead. The adjective "subconscious" means "$subaware_1$"; neural activities that do not send their signals across the

1 _Receptors and Sensory Perception_, 1955.
2 See, e.g., Smart, _op. cit._, pp. 102-3.

awareness line are subconscious activities; we analyze our visual signals subconsciously, we regulate our digestive processes subconsciously, and perhaps the great decisions of our lives are influenced strongly by subconscious activity. (See §§38 and 41.) The speech centre (everything on the speaking side of the awareness line) has direct access to a very limited part of the brain, viz., the output signals of the rest of the brain. For the awareness line, only direct access counts as access. We are not aware of the contents of signals that converge to produce the output. We are only aware of the content of the output. Epistemologists of all stamps would agree that the only awareness is direct awareness, however this is to be interpreted. Then, if the psychoanalytic region, the Unconscious, is treated not as the opposite or complement of consciousness$_1$, but of awareness$_1$, the difficulties about areas and fringes disappear. On this view, the Unconscious could be unmetaphorically identified as an area of the brain - as some part of it on the far (inner) side of the awareness line. No area need be posited to house consciousness$_1$, which is nothing more than the state of being awake, or awareness$_1$. The geometrical guide for awareness$_1$ is a line, not an area or volume.

Several corollaries of importance can be drawn from this treatment of "aware" and "conscious". Animals are conscious$_1$ but not aware$_1$. People can be aware$_1$ of things while asleep, while only partly conscious$_1$. This may seem strange, but there is no reason why in dreaming, signals should not cross the awareness line; they certainly do when one talks in his sleep. And do we not say, and mean in quite the right sense, that in our dreams we were aware of this and that? "Suddenly I was aware of a man with a knife, and that scared me so much I woke up." Any machine with servo-mechanisms that duplicated the functions of the nervous system to a reasonable degree could be said to be conscious$_2$ - and conscious$_1$ when turned on. A machine like the perceiving machine would be both conscious$_2$ and capable of awareness$_1$.

A popular philosophical question these days is: What would a machine have to be to be conscious?[1] The question cannot be answered, of course, until it is decided just what consciousness amounts to. Once this is decided, the question is answered. If the notion of consciousness is piled with excess baggage, including, say, the ability to love, to enjoy strawberries and cream, to be lazy, to appreciate Bach - in short, if the notion is made indistinguishable from the notion of <u>mature humanity</u>, then the

1. See, e.g., Danto, "Consciousness and Machines" in Hook, <u>op. cit.</u>, and Anderson (ed.), <u>Minds and Machines</u>.

machine that fills the bill will have to be simply a man-made man. This is not impossible in principle - though certainly in practice - but if the machine were built, it would be a lot more than just conscious. A requirement discussed by Scriven is that anything conscious must be alive. But, as he points out, anything that could be correctly called alive could not correctly be called a machine (at least on the most obvious criteria of life), so the requirement is necessarily impossible to fulfill. Scriven thus rejects this condition, but he says for other reasons that "machines will never be conscious, because we have come to see that a reproduction of a man sufficiently exact to be conscious is too exact to be still a machine."[1] He admits that this may change. If the word "conscious" came to shift in meaning due to new theories, then machines might properly be called conscious. It is just such a shift (for theoretical purposes at least) that I am proposing.

Having completed this preliminary analysis of awareness and consciousness (images have yet to be exorcized, and thinking as a creative act must be described), I can attempt an answer to a question raised in the first two chapters: Are introspective reports in fact referential?

---

1 Op. cit., in Hook, op. cit., p. 36.

The question may (hopefully) have lost a bit of its urgency in the meantime. I have argued that there are no objects that are reported or described when one makes what is called an introspective report. There is nothing well-referred to by "thought" in "I have a thought ... ", by "idea" in "I have an idea ... " and so forth. There are, however, neural signals, and these may be ascribed content. We do not intuit or perceive this content (if these words are meant to refer to some actual activity or process); all that is the case is that a signal of a certain content crossing the awareness line produces a certain utterance provided the neural chain is not inhibited or otherwise tampered with on the way to the muscles. This does not mean that the utterances considered as statements are not _meant_ as referential. After all, perhaps a misguided person can mean to refer to something with the word "sake" in "I'm doing it for Fred's sake"; perhaps statements about voices are meant as referential. _Are_ statements about voices really referential? If the question can still be asked, it is certainly not an ontological question, and the same holds for the question about introspective reports. My purpose in holding the question in abeyance was to avoid prejudging the ontological questions. Having avoided this, it makes no difference to the case for physicalism how the purely linguistic or logical issues are decided. The notion of meaning something by saying something is separable from the questions concerning how human beings produce the utterances they do.

This indifference to "the actual meaning of the words" is part of a bigger indifference about ordinary language – for the purposes of propounding physicalism. Once the ontological questions have been settled, and the puzzles about awareness and thoughts dissolved, one can talk however one wants, provided only that one does not lapse back into discarded theory. In practical, i.e., non-theoretical, contexts it does no harm to say "I have a thought", or "I am intuiting an apple-shaped sense-datum", or "My transcendental ego is being appeared to redly and roundly". The context is a practical one if one's audience draws no inferences other than those expressed by such responses as "Oh, you've finally noticed the apple, have you?".

In line with these remarks, I shall drop the subscript from "aware_1" and "conscious_1" in the remaining chapters, and use "aware" and "conscious" in the non-ordinary senses I have delineated. I shall also relax the ban on such idioms as "thinking a thought" and "reporting an experience", reverting to more rigorous expressions only when my interpretations of these idioms may seem to be in jeopardy.

CHAPTER 9.

Perception and Sensation.

24. The Physical Features of Perception.

In Chapter 4 an undetailed account was given of the neural organizations that could produce the phenomenon of sight. In this chapter a more detailed description will be given of the physiology of vision, with the aim of drawing more specific philosophical conclusions. Following the example of previous chapters, the account will depend on the definition of non-ordinary counterparts for a few key words.

The ordinary, or at least semi-ordinary, word "sensation" is the first candidate for redefinition. Inherently weak, this word has been crippled by philosophical pressure. The reification in ordinary usage of strange events in the nervous system ("I have a crawling sensation in my back", "There is a tingling sensation in my foot") has been adopted and expanded by philosophers to include the objects of visual perception (visual sensations or sense-data) and such things as sounds, smells, and pains. Then these objects have been located, variously, in the mind, in the brain, in the eye (for visual sensations) or in the foot (for pains of the foot), in corresponding parts of the "phantom-body", on the near outside surfaces of physical objects, and nowhere. The initial step in clearing up this morass is to define sensations as impulses (or, with an increment of interpretation, signals) carried by the afferent network of the nervous system. This redefinition does considerable violence to the ordinary meaning of the word, since admittedly the man in the street can speak of his tingling sensations without a scrap of knowledge about the nervous system. But the definition does more or less follow the usage of the medical profession, where one speaks of sensation returning to an injured limb, meaning by "sensation" (or at least referring to by "sensation") the impulses that are once again travelling to the brain through the afferent fibres. Later I shall propose a substitute expression for the use of the man in the street with theoretical scruples - for the philosopher, in other words.

Under this definition sensations are things the analysis of which will be aided by microscopes, dissection, and so on. Sensations are not private, but simply in a body, more readily observable and accessible to the physician, in fact, than to their owner. The questions that can be asked about these sensations are what Ryle calls "wires and pulleys" questions.

We are not aware of our sensations. We are aware of the content of some of our sensations - of those that cross the awareness line. If I observe my own impulses (with the aid of surgery, micro-electrodes, oscilloscopes, mirrors, etc.) then I may be aware of the content of some sensations to the effect that I have certain sensations going on, and in that sense I might be said to be aware of my sensations. This would not differ in kind from being aware of my hat (aware of the content of some signals that my hat was....). Without wishing to lean too hard on a grammatical distinction, I suggest that awareness is best seen as awareness that some state of affairs is such-and-such, rather than as awareness of some object. A heuristic (only) analogy is that just as the radio officer on a ship is not alarmed by the beeps of Morse Code he hears, but by the message they transmit, one is not aware of sensations, but of the information or content they transmit.

To rephrase Putnam's point,[1] it is necessary to have sensations, but it is not necessary to be aware of them in order to have certain kinds of knowledge - in order to see. On a sympathetic interpretation, Husserl's material or hyletic phases are just these sensations, for hyletic phases are the temporal (non-eternal), real, informing part of perception of which one is not conscious.[2] The

1 Op. cit., p. 155.
2 See his Ideas.

contribution of these signals is limited by their nature as signals and by the capacities of the organs that produce them. Signals are finite and thus carry a finite amount of information, however many convergences have occurred in their production. However much can be said about the subject matter of a signal's content (the physical object, say, that reflects the light rays into the eye), only a part of this can be "said" in the signal's message. For each sense organ there are thresholds of discrimination, and since distinctions below these thresholds are not registered by the organs, information about such distinctions cannot be transmitted by the signals produced. These points are obvious, but a philosophical problem that hinges on them is far from obvious.

I have already described the analysis by signal convergence that occurs for each sense. Here I will go into somewhat more detail about the organization of visual signals, since vision is the most important and most complex sense. The results of the examination can then be applied to explanations of the other senses.

The mosaic of stimulation on the retina is not transmitted intact to the cortex. The pictorial, as opposed to the reportorial, element of vision is broken up right at the retina. The visual signals sent to the brain carry information about motion, colour areas, edges, verticals and horizontals, and there is no evidence that these signals are somehow decoded at some point, thus reintroducing the pictorial element. Television receivers decode their signals and reproduce the pictorial element; brains do not.

These sensations, whose subject matter is not points of light on the retina but shapes and motions and positions on the retina, are collated by neural organizations set up by past experience (past sensations) and present "interests". That is, in addition to the pathways that past patterns of stimulation have branded on the brain, particular convergences can be temporarily encouraged through synaptic facilitation, if some particular pattern of sensation is sought as a guide to present behaviour. This is the phenomenon called "set" by psychologists. For example, there is in most experienced human beings neural machinery for interpreting a wide variety of stimulus patterns as indicating the presence of deer. In the hunter who has spent eight fruitless hours looking for deer, these particular convergence-possibilities are temporarily facilitated and discrimination requirements are relaxed. This puts the deer-signalling neuronal combinations on a hair trigger, with the dangerous result that when almost anything moves the hunter <u>sees</u> a deer. The ability to set seems to depend on previous relevant experience and hence on the strength of the existing neural organizations, or in other words their well-established identity. Radio operators on ships can set for two complex signals in Morse Code: their own "call signal" and "Mayday", the international distress signal. They can then sleep, undisturbed by the cacophony blaring in their ears, but if either of the two signals set for occurs they wake up immediately. Recent experiments in scanning lists of words for wanted words (and even for wanted types of words, such as names of animals) suggest that the particular convergences sought can be facilitated, and all unwanted convergences inhibited to the point where nothing or next to nothing in the list except the sought words can be reported or remembered. Subjects claim they do not "read" the individual words at all.[1]

In addition to these temporary facilitations there are more or less permanent facilitations (or relaxed discrimination requirements) by which the actual retinal data are warped. The best example is the Fuchsian completion. If a subject is shown a circle from which a tiny bit of arc has been erased he will report seeing a complete circle (unless he is on the lookout for gaps), even though the size of the

---

1 Ulric Neisser, "Visual Search", in Scientific American, June, 1964.

gap is well within the power of resolution of the eyes. In this case the convergence requirements for the firing of circle-informing signals must be slightly substandard.

It may seem that the complexity of the distinctions involved in sorting out the relevant patterns in this way is just too great to allow the human neural net to accomplish the task. This problem of "the recognition of universals" figures often in psychological and cybernetic research. How can a neural net distinguish chords regardless of pitch, shapes regardless of orientation, individual human voices, and letters of the alphabet camouflaged in a tangle of other lines? Neural net analyzers have been made that can distinguish letters of the alphabet provided the variations of size, orientation, and shape are within certain narrow limits, and devices for recognizing handwriting and voices are being developed. It must be remembered that human discrimination is not infallible and does depend on certain favourable conditions. As Arbib points out, "We tend to bring any object that attracts our attention into standard position and orientation so that the visual trace of it formed in our nervous system varies within as small a range as possible."[1]

---

1 Brains, <u>Machines and Mathematics</u>, p. 108.

The immensity of the problem is diminished somewhat by the fact that we cannot readily or without practice read words that are upside down or placed so that their image falls on the periphery of the retina; we cannot usually recognize a picture of a friend if it is shown to us upside down; we cannot describe objects placed at the periphery of the visual field.[1] It must also be remembered that the human brain has approximately 10^{10} neurones with which to do its work, a figure which dwarfs all present or even

---

1 There is a rare form of abnormal perception, simultanagnosia, which vividly points up the necessity of well organized eye-motion controls based on low-level retinal signals. In simultanagnosia, the subject is able to identify and describe parts of his visual experience, but is unable to put the parts together to form significant wholes. For example, shown a picture of a boy standing on the roadside beside a broken bicycle hailing a passing car, the subject reports: "There is a boy - on a bicycle - a car - somebody is waving - there are trees." This is not tunnel vision or a similar narrowing of visual scope, and the subject is well aware that he is simply unable to put visual details together. (Kinsbourne and Warrington: "A disorder of simultaneous form perception", Brain, 1962.) Part of the difficulty is failure to coordinate eye motion in scanning the picture, but this seems to be due not to muscle or nerve failure, but to higher organization. Luria, Pravdina-Vinarskaya, and Yarbus (in "Disorders of ocular movement in a case of simultanagnosia", Brain, 1963) use an ingenious apparatus that records the motion of the fixation point on a graph. They showed that in simultanagnosia thefixation point followed an erratic course unrelated to the drawing being viewed, whereas the graph of a normal viewer's eye motion virtually reproduced the drawing to the point where facial features in the drawing were perfectly recognizable on the graph.

practicable future computers. When it is added that that part of the brain which analyzes the afferent barrage is also known to have a more pliant and adaptive structure than computers have, the situation no longer seems impossible or even improbable.[1] Furthermore, if the brain does not perform this function, what does? The brain is surely the right sort of thing to do the job.

One of the functions of the analyzing mechanisms in the brain is the discrimination of poorly lit white objects, brightly lit black objects, and blue objects at sunset. The ability to do this must depend on previous experience, which establishes the appropriate analyzing organizations. Then it is not surprising that when objects are placed in extraordinary light we can be fooled about their colours; we simply have acquired no neural mechanism for sorting out these cases, and they are thus wrongly analyzed by whatever mechanism we have.

Experimental corroborations of this view are numerous and interesting. Van Senden describes the case of congenitally blind adults who are given sight through surgical operations.[2] At first the patients announce that the

---

1 Even if the part of the brain that has this function were in the order of one one-hundredth of the brain, 10^8 neurones (approximately 1/100th of the total) is still much greater than the number of analogous modules in modern computers.

2 Van Senden, Space and Sight, 1960.

experience is simply painful, and often they say that all they see are flashes of light and motion. They cannot identify the simplest objects under optimum conditions. One should take their descriptions of flashes of light with a grain of salt, of course; since they have never before had the opportunity to call anything first hand a flash of light, there is no reason to suppose that this first description is at all trustworthy. Before one can trust their accounts of what they see, they must have shown that they can correctly describe actual objects, things everyone can see. Otherwise there is no reason to believe and every reason not to believe that they have learned the correct use of visual terms.[1] A safe conclusion is that their visual sensations are disorganized and hence fundamentally content-less to them. It takes as long as a month before the patient can identify simple brightly coloured objects, and still longer for him to identify such objects under different conditions of position, distance, and lighting.[2] A truly useful sense of sight takes these patients years to

---

1 Their using words incorrectly does not amount to their misidentification of any neural signals or their misdescription of any sense data or other phenomenal objects. They are misdescribing what is to be seen in the outside world - because they do not yet have the equipment for correct description.

2 E.g., the patient can identify a red golf ball placed on a table, but not if it is hung from a string.

develop, and many patients unfortunately give up, fall back on their old reliance on touch and sound, and never learn to use their eyes.

The proposed explanation of this is that the patients, never having had any visual signals to coordinate, have developed the tactile and auditory signals to such a degree of discrimination that the new flood of signals serves for them no vital function, and since the brain has been programmed independently of these signals, they are treated as mere interference.[1] This would explain why these patients take so much longer to learn to use their eyes than infants do; for the infant, all afferent signals are on a par as far as utility is concerned. The analysis of the visual signals depends strongly on the simultaneous convergence of tactile and motor signals, as is borne out clearly by new-sighted patients, who can make no headway whatever in using their eyes until they have had the opportunity to touch and look at the same time.

A similar result was obtained by Kohler in 1951, using eyeglasses that distorted or inverted the retinal image.[2]

---

1 The extent to which there may also be atrophy or damagé in the neural pathways has not yet been determined.

2 The experiment was first performed by Stratton in 1896. For an account of Kohler's experiments, see Granit, <u>op. cit.</u>, p. 399 ff.

In Kohler's experiments subjects wore glasses with an inverting mirror or distorting lenses continuously for as many as 124 days. With the inverting glasses everything seemed upside-down to the subject at first, and any activity - walking, writing, maintaining balance - was next to impossible. After a few days, provided the subject's eyes were trained on his own body and immediate surroundings, normal or right-side-up vision was reported. Glancing off, however, re-inverted the subject's vision. After a period of days in which vision could be flicked back and forth like an illusion drawing, completely right-side-up vision resulted, of such reliability that subjects were able to ride bicycles and even to ski. When the glasses were removed, vision was upside-down for several days before reverting to normal. The wearers of distorting lenses at first were unable to recognize faces and objects, and unable to read or write. But again, after a week or so they had normal vision, and after removing the lenses experienced distortion for a few days. Furthermore, with the distorting lenses colour vision was at first impaired by rainbow effects, which gradually disappeared, only to reappear for a few days on removing the lenses. These results fit nicely with the view that signals arriving at the brain do not mean anything intrinsically, and that discrimination and recognition are all a matter of coordination of visual with other afferent input.[1]

25. The Reportorial Nature of Sight.

As suggested earlier, there is no one level of neuronal interaction that can be specified as the turning point from afferent to efferent signals. The awareness line can be considered a turning point perhaps for vocal efferents (and partially for writing efferents), but there is no one place where all afferent signals arrive, are judged, and thus trigger behaviour. At the very lowest levels, afferents from the retina trigger signals controlling eye movement, focus, and blinking; and most reflexes, such as ducking, do not depend on a high-level analysis of the input. The convergence of visual signals with tactile and proprioceptive signals does not occur at the highest level of analysis, for the highest levels in fact depend on the previous convergences of these signals. Behaviour does not seem to be initiated and controlled from one level in the

1 Kohler draws other conclusions from his results, mainly concerning the existence of phenomenal space, etc., with which I do not agree. An interesting sidelight is that some animals - at least chicks - cannot adapt to prismatic lenses, which suggests that their afferent-efferent connections are more permanently established. See Hess, "Space Perception in the Chick", Scientific American, 1956.

analyzing mechanisms, and thus it would be unwise to suppose that there is anything like a central switching-yard, a brain within the brain, where input is turned into output.

For the purposes of a very limited and expressly metaphorical explanation, however, it will be convenient to consider the highest-level visual signals as arriving at Visual Headquarters, bearing in mind that such a place might be widely scattered and only arbitrarily determinable. I wish to use the device of Visual Headquarters in an extended and completely fanciful allegory designed to put across a certain point about sight which is difficult to put across otherwise. I am driven to this unrigorous course by the existence of deep-seated and pervasive misconceptions about sight that exist only in highly metaphorical form, and thus are practically invulnerable to straightforward demolition.

The information arriving at Visual Headquarters is like the information arriving on a newspaper's teletype machines. It carries the date-line of the fixation point, but along the way it is edited and interpreted. The final product is a strictly limited account of what at the moment is "judged" (by the editorial organization along the way - the mechanisms of afferent analysis) to be going on at the fixation point in the outside world (not what is going on on the retina).

The reporters in the field, the retinal receptors, do not send snapshots or movies to Visual Headquarters; what arrives at V.H. is not reconstituted like a wirephoto into an image. The receptors send isolated non-pictorial reports which are then compiled and edited (by convergence) on their way to Visual Headquarters.

As the roving reporters are called from place to place (as the eye moves and the fixation point shifts) - and this may be in response to "local" (retinal) supervisors or on orders from V.H. - the reports at V.H. pile up, with old ones being discarded as new ones come in. At any moment there is a backlog which, if put together, "paints a picture" of what is going on out in the field - not in the sense that the reports could be flashed on a screen to form a spatial picture, but in the sense in which an historian can "paint a picture" of Ancient Rome. In physiological terms, the high level signals trail off gradually from the instant of firing, so that at any moment there is a collection (however scattered) of high level visual signals which can jointly contribute to further neural activity and hence behaviour.

The point is that Visual Headquarters can operate as a control centre without the benefit of pictures, maps, movies, or other images. If there were such images there, they would simply have to be seen and interpreted and reported, so there would be no gain. There are no homunculi or other things to look at images, so images would not help. The function of sight is to get information into the form in which it can contribute to control. Aerial reconnaissance in wartime is of no use in contributing to control if no one interprets the pictures. What goes on at Visual Headquarters is like a picture in about the same way descriptive writing is like pictures.

The difficulty with this allegory is that there is also nothing at Visual Headquarters to read or listen to non-pictorial reports. The difference between neural impulses and messages proper is that messages proper must be read and understood if they are to be effective. Neural impulses are like messages in the sophistication of their function; because they have such sophisticated functions they can be ascribed content. But they are also like the electric-eye impulse running from eye to door-opening device; the door-opener does not read the impulse. Thus neural signals are neither fish nor fowl, but they do have company in their uncertain status. The "orders" radioed to satellites and rockets are not really orders, since if the satellite is in working order the result is obtained just as in the door-opener case. Perhaps one can say that a soldier is in "proper working order" if he carries out commands just like the satellite. Is the soldier well-programmed? Or does the satellite understand? A good soldier obeys orders because he is loyal, or well-trained, or stupid-and-dogged, or machine-like, and in any event because he <u>understands</u> the orders. But cannot all this be seen as somehow a matter of how he is programmed? The point is that drawing a fine conceptual line between ordering and button-pushing is an enormous and fruitless task. Orders to satellites are perfectly well understood things whatever one wants to call them, and I am saying that neural messages are the same sort of things.

Some of the outputs of Visual Headquarters, which, it must be stressed again, is composed of parts related only in their approximate function and probably not at all in topographical contiguity, go to the speech centre, and hence cross the awareness line. However, one should not imagine a coaxial cable from V.H. to the speech centre; V.H. is not necessarily a closed area from which such a cable could run. Some high level visual signals (depending on what is appropriate at the time) cross the awareness line sometimes, and that is all. Other continuations contribute to the control of this or that non-verbal behaviour and so forth. If the synaptic chain for a certain message is inhibited at some point prior to the awareness line, then one is not aware of that jot of information, even if one's eyes are properly aimed, open, normal, functioning, and so forth. When one stares at a spot on the wall, but is "lost in thought", one is not aware of the spot on the wall, even though the fixation point signals may continue to be facilitated. Their continuations are blocked later.[1]

This analysis of sight leads naturally to a redefinition of the word "see" parallel to the definition of "aware". One $sees_1$ <u>that</u> p, when p is the content of a (predominantly) visual signal, and this signal crosses the awareness line. Animals see_2 that p where p is the content of some visual signal that contributes to behaviour somehow. Then people see_2 that p, but do not see_1 that p, when the signal contributes to the controls of some activity, but does not cross the awareness line. The automobile driver $sees_2$ all the corners, poles, etc., that he does not also see_1.

This usage fits part of ordinary usage quite well: we do say that we see <u>that</u> such and such is the case. This

---

1 Cf. Sprague, "The Mind-Brain Problem", in Hook, <u>op. cit.</u>, p. 69: "But the physiology of vision is beside the point, when I cannot see my collar button although I am looking straight at it." It all depends, of course, on where one draws the line for the physiology of vision. Nothing in the <u>eye</u> may be different, but if Sprague is suggesting that there is <u>no</u> physiological difference between seeing and not seeing the collar button, he is mistaken - even though no one can say right now just what the difference is.

partial agreement is a trivial matter. We might not have had the idiom, or there might be languages in which the idiom translated "to see that p" was deviant, but that would not affect the desirability of the theoretical use. And hence the partial disagreement of usage is trivial. We do say we see objects, but what of it? To someone who argues: "But when I see an apple, there is an object, there really is some thing I see" (this sentence must be said with great emphasis on each word to achieve the proper mystifying effect), I reply: certainly, and it is the same object one touches and eats and is nourished by. That is what the ordinary idiom means. Certainly it is true that when one sees an apple, one must necessarily see that it is where it is - at least in the sense that it is front and centre and about arm's length away from the eyes. As in the case of "aware", I do not want to lean on this grammatical matter. I merely urge that the "see that p" idiom is safer since it does not even suggest the introduction of strange objects.

There is quite a different, strictly relational, meaning of "see" pointed out by Warnock: " ... the truth of the statement that Jones saw a fox is not impugned by Jones' misidentification of what he saw nor by his failure to identify or mis-identify it ... ".[1] In this sense "see" relates two objects, seer and seen, but as Cyril Burt points out, this use is only for third person and first person past tense ascriptions ("he sees" and "I saw"). For first person use ("I see"), the word is applied on the basis of particular descriptions.[2] Philosophical timidity in the face of this patent inconsistency between "he sees" and "I see" has led to the volumes of devious descriptions of objects to which the reporter may be related when he says "I see": sense-data, sensa, noemata, ostensible physical objects, and other queer entities.

26. Sight without Images.

Shorter, in his excellent article, "Imagination"[3], describes imagining as more like depicting than like painting a picture. We can imagine something without going into great detail. If I imagine a tall man with a wooden leg, I need not also have imagined him as having hair of a certain colour, dressed in any particular clothes, having or not having a hat. If, on the other hand, I were to draw a picture of this man, I would have to go into the details. I can make

1 "Seeing", Proc. Aristotelian Society, 1954-5, p. 205.
2 "The Concept of Consciousness", British Journal of Psychology, 1962.
3 Mind, 1952.

the picture vague, but the man in the picture must either be supposed to have a hat on or not. As Shorter points out, my not going into details about hair colour in my imagining does not mean that his hair is coloured "vague" in my imagining; his hair is simply not "mentioned" in my depiction at all. If I write down a description of a person it would be absurd for anyone to demand that my description cannot fail to mention the colour of his necktie. Similarly it would be absurd to insist that one's imagining of someone must go into the colour of his necktie. The point that Shorter misses is that <u>all</u> "mental imagery" is like depiction in this way. Not only imagining and dreaming, but seeing and hallucinating as well. This is all just part of the obvious truth that there is no projecting screen, no theatre, in the brain.

Consider the difference between a bit of descriptive writing and a motion picture. Let us take one of Tolstoy's battle scenes from <u>War and Peace</u> and the film version of the same scene. The film version goes into immense detail, if one cares to examine it inch by inch, whereas the "picture" painted by Tolstoy simply does not go into a lot of the detail that the film cannot help but go into (such details as the colours of the eyes of each soldier). And yet Tolstoy's description is wonderfully vivid. My point is that seeing is more like the written Tolstoy than like the film. The film is not like seeing (although seeing the film is like seeing the outside world), but like the image on the retina; it is the raw material of the seeing process, not its end product. Seeing is like reading a novel at breakneck speed - written to order at breakneck speed.

Of course the analogy should not be carried too far. There is nothing in sight neatly corresponding to the written word; impulses are not, in spite of some neurologists' views, analogues of the beeps of Morse Code. There is no sensation language in any sense as strong as this. And again, nothing or no one <u>reads</u> the neural messages, although if they cross the awareness line, they lead to expressions in words.

It is easy to see how Shorter missed the point about seeing being just like imagining in being like depicting. Seeing, like reporting, is to some extent dictated by the facts, whereas imagining is writing pure fiction. While the fiction writer can make up just what facts he wishes, and is not bound to go into infinite detail, the reporter cannot write down all the facts, but must send in a limited, edited account. There are two "introspectively observable" limitations on imagining that can now be explained. The first is that all imagining is from a point of view.

The physiological description of imagining proposed is that in imagining, the actual afferent barrage is inhibited, and high-level visual signals that would occur if one were seeing what one is imagining are internally or reflexively stimulated. The success of such an operation would of course depend on the "strength of character" of the neurone patterns involved. Until one has developed some firing patterns to have the content of, say, seeing a red object and seeing a golf ball, one cannot imagine a red golf ball. Hybrid signals can of course be produced, but lacking the components, one cannot make hybrids.

If this is in general outline the mechanics of imagining, it follows that whatever scene or object I can conjure up must be presented at a particular angle to the "imagining eye". Imagining is <u>as if</u> one were seeing, and seeing is always from a point of view. If I imagine a house I must imagine the front of it (with the observing eye stationed in front of the house) or the back of it, or the top from a bird's eye view. If I imagine a room I (or at least my point of view) must be either inside the room or outside it.[1]

---

1 A critic of mine once argued that he could imagine a spherical room of mirrors without imagining his own reflection - but of course his <u>point of view</u> was either inside the room or outside the room.

The second limitation is that I can construct in my imagination objects of only limited complexity. If I wish to imagine a monster with seven different kinds of legs, wearing a patchwork coat, I must build up the image piecemeal; I cannot "see" the monster in all its complexity at a glance. But if there were such a monster I could not really see it in all its complexity in a glance, but would have to run my eyes over it. The difference is that in actual seeing the details already glanced at are there to come back to if I forget. I need not carry them all in my memory.[1] It might be this that Hume meant by saying that imagining was less vivid than seeing. In any other sense imagining can be extremely vivid.[2]

No doubt the primary reason we feel that seeing is genuinely pictorial is because whenever we examine our own experience of seeing, whenever we set out to discover what we can say about what we are seeing, we find all the details

---

1 In the phenomenon of "eidetic imagery", the details remain as in sight, and can be read off by the subject. See Gordon Allport, "Eidetic Imagery", in the British Journal of Psychology, Vol. 15, 1924.

2 Can animals imagine? They cannot imagine as people do, since any "as if" signals that were stimulated by the brain could not cross the awareness line - because there is no awareness line. But there is no reason then to suppose they cannot dream (if dreaming is just a form of as-ifing). This would not mean their imagining or dreaming would be blind somehow, any more than their seeing is blind. They simply cannot report it.

we think of looking for. When we read a novel questions can come up that are not answered in the description, but when we are looking at something, as soon as questions come up they are answered by new information as a result of a shift in facilitated signals or fixation point. The reports of perception are written to order; whatever detail interests us is immediately brought into focus and reported on. When this occurs, one is not scanning some stable sense-datum; one is scanning the outside world, quite literally. It is absurd to suppose that one can scan one's scannings.

One can no more become interested in a part of one's visual experience without bringing the relevant information to the fore than one can run away from one's shadow. The rule is: seek and ye shall find. It is this characteristic of seeing that is no doubt vaguely recognized by those who argue that introspection disturbs the process under observation. The concept of introspection thus seen adds one more absurd level to the "inner eye" fallacy. For when we "introspect", there is an inner eye watching the watchings of another inner eye observing the sense-datum - and all this watching infallible! Introspection, if anything at all deserves the name, is merely saying what one can say, uttering the messages that cross the awareness line, and then examining the utterances, not the "experience".[1]

Since the only limits to the colour-spatial information we can receive from our eyes are discrimination thresholds, it is tempting to suppose that all the details we can know about are always there when we see something. There is no limit to the description one could write of a painting; one could describe it inch by inch and centimetre by centimetre, the wide-angle view and the detailed view. Such a description could be very nearly complete at any time after, say, the hundredth page of description, but the description would still not be a painting. And there would still be things about the painting that the description said nothing about.

Nevertheless, the notion of imagery and phenomenal space must, I fear, die hard. Is there not a real difference between actual physical space and "subjective" space? In real space things are the size they are: they do not diminish as they get farther away. Parallel lines do not converge at the horizon in real space. But this is a matter of perspective, and perspective is a matter of optics and simple geometry, and is a phenomenon in real space. A camera "sees in perspective", but it has no private subjective space. The image on the human retina is in perspective;

1 Cf. Mace, op. cit.

it is the final image in the seeing process, and is almost immediately transformed into information about size and distance.

It is true of course that when we see, we do not simply see that there is a table in front of us, but a table of a particular colour and shape in a particular position and so forth. All this means is that the information we receive is vivid and rich in detail. It is pictorial in the way good writing is pictorial. It is, as we say, graphic.[1] This is not true of the vision of many lower animals. The frog, for example, can see that there is a small moving object before him, but he cannot see that it is a fly or a bit of paper on a string. If the small object is not moving, he cannot see it at all, because motion signals are required for the higher level synapses. A frog left in a cage with freshly killed (unmoving) flies will starve to death, because it has no neural organization for sending the signal: there is a fly (moving or still). Dangle a dead fly on a string and the frog will eat it and survive. There is no physiological evidence (and what other kind of evidence could there be?) to suggest that human sight differs

---

1 Calling writing graphic or pictorial depends on a prior concept of images, or pictures, but does not depend on there being mental images.

in kind from frogs' sight, although there is no doubt that it differs greatly in degree.[1]

Vividness and complexity are not sufficient reasons for positing images. Trained musicians can read difficult scores, and appreciate and understand them perfectly - one might even say they enjoy them - and this is accomplished without singing or humming or even humming to oneself. If there is no reason to posit some psychic or cerebral orchestra playing the music in the case of score reading, why should one posit image formation for the appreciation and understanding of the visual world?

It may seem that I am beating a dead horse. No one comes out and says in so many words that there are ghostly images inhabiting some never-never land of the brain. But the notion of images is behind a great deal of otherwise very sophisticated philosophy. Once images have been banished, the problems encountered in these philosophies disappear.

Consider the perennial philosophical problem of the Tiger and the Stripes. I can dream, imagine, or see a striped tiger, but must the tiger have a particular number of stripes? If what I am seeing is an object, then as an

1 Muntz, op. cit.

object it must have a certain number of stripes, and one should be able to pin this down with such questions as "more than twenty?", "less than thirty?". But if what is seen is descriptional, i.e., that there is a striped tiger out there, the question cannot be asked. Unlike a snapshot of the tiger, one's seeing need not go into the number of stripes at all.

Wittgenstein's duck-rabbit is another case in point. What is the difference between seeing it as a duck and seeing it as a rabbit? The image (on the paper) does not change. But there could be more than one description of that image, depending on which of two neural pathways happened to be facilitated.

Russell Brain reports that patients recovering from eye or brain injuries sometimes report that as their sight returns they are first aware only of motion.[1] Not motion of light patches or objects, but just motion. This sort of introspective report could not possibly be true or even sincere on any imagistic theory. Motion, all by itself with no vehicle, is not something of which there can be an image. But there could be a neural report of just motion. One of Kohler's subjects in the inverting spectacles

1 "Some Reflections on Brain and Mind", in Brain, 1963, p. 381.

experiments described his experience when shown two human heads, one upside-down, both peeping out of a hole in a background, by saying that both heads were both right-side-up and upside-down. Kohler (in the film on his experiments) drew particular attention to the fact that this state-of-affairs is not graphically representable. He did not go on to conclude that since no "mental" image could represent it either, so much the worse for mental images.

Of all the problems that have led philosophers to posit phenomenal space and phenomenal objects, the most tenacious has been the problem of illusions and hallucinations. It need hardly be mentioned that the descriptional theory of experience dissolves this problem. There can be little doubt that hallucinations are caused by freak neuronal discharges. Stimulation by electrode of micro-areas on the visual cortex produces specific and repeatable hallucinations.[1] When one experiences an illusion or hallucination

---

1 Penfield, The Excitable Cortex in Conscious Man, 1958. It would be expected that hallucinations, like imagination, would depend on the pre-established character of the neuronal patterns involved, and this is in fact supported by evidence. Amputees usually experience "phantom limb" sensations that seem to come from the missing limb; a legless man may feel that he not only still has the leg, but that it is itching or hot or bent at the knee. These phenomena, which occur off and on for years, are nearly universal in amputees, with one interesting exception. In cases where the amputation occurred in infancy, before the child

the only thing that there is that is not present normally is the non-veridical description carried by a physical signal. And where is the description, and how much space does it take up? It is in the brain, being carried along by electro-chemical impulses, and it takes up, I suppose, as much space as the neurones do that are transmitting it, just as my description of perception takes up a certain amount of space on pieces of paper.

Whatever the reasons for positing images or other ghostly objects, there are strong reasons against positing them. Ryle's embarrassing questions are intended to reveal just how strange these objects are: Where are sense-data? What are the dimensions of phenomenal space? How long is a polar-bear sense-datum? It may be argued that similar absurd questions can be asked about other objects that are not so strange. What colour is an electron? What is the temperature of a neutron? How heavy is a lacuna? But there are perfectly straightforward ways of answering these questions. Temperature and colour are macro-properties; they have no application to micro-entities. The word "lacuna"

---

developed the use and coordination of the limb, phantom limb is rarely experienced, and in cases where amputation occurred just after birth, no phantom limb is ever experienced. (Simmel, "Phantom experiences following amputation in childhood", Journal of Neurology, Neurosurgery and Psychiatry, 1962.)

is a handy short-cut for longer expressions about destroyed parts of sentences. But how is one to explain an image to which no yardstick may be set? How does one justify a space that takes up no space and yet is within another space? Perhaps this talk of images is metaphorical. But if it is just metaphorical, then there is no need to espouse it in the face of mounting criticism. What sense can be made of saying that certain facts _must_ be expressed in a certain metaphor? Can there be facts that are only metaphorical, in the sense that one can express them only metaphorically?

Consider the status of a character in a novel. Where is Tom Jones? On the pages of the book? In Fielding's brain? Or mind? How does he stay alive? It is no good explaining that mental objects are, like Tom Jones, fictions, but _useful_ fictions, because for theoretical purposes mental objects are not useful. They clutter up our conceptual scheme with strange spaces with stranger inhabitants, they generate the unnecessary problem of certainty of description, and they inadvertently baffle and sidetrack otherwise talented neurologists.

Thus Smythies is led to distinguish symbolic from non-symbolic representative mechanisms: telegraphy and writing are symbolic, paintings and television are non-symbolic.

Then, since perception is non-symbolic (like television!), we should not call sense-data symbols of outer reality.[1] It is perfectly true that perception is not in a language or code, but Smythies does not stop there. Mental images have him so much in thrall that he then goes on to hunt for spatial patterns in the brain "isomorphic" to the objects perceived.[2] And, if I understand him correctly, he is searching the brain for a television-type scanning system.[3] Smythies makes the traditional philosophical point that we do not _see_ sense-data - we _have_ them, but then lapses into defeatism: "The brain, in fact, as a machine simply cannot construct the sense-data that, as we have seen, play their part in every perception."[4] This leads him to toy with the idea that neurological explanation must finally depend on one non-physical joker, the Pure Ego.[5] The Pure Ego, for Smythies, will be the ghostly eye that infallibly observes (with privileged access) the mental objects.

---

1 "The Problem of Perception", _British Journal of the Philosophy of Science_, 1960, p. 227.
2 _Ibid._, p. 234, and _Analysis of Perception_, p. 18.
3 _Analysis of Perception_, p. 68; "Analysis of projection", _British Journal of the Philosophy of Science_, 1954, p. 126. He is looking for Smart's and Feigl's checking mechanism, and expects it to be something like a television camera aimed at a television screen.
4 _Analysis of Perception_, p. 18.
5 "The Problem of Perception", p. 228.

A common concept in neurological circles is "projection", the idea that the brain somehow projects its sense-data into some sort of outside space, possibly into ordinary physical space. (Why outside? Is it because it is too dark inside the brain to see?) Although this idea is often discredited, Smythies' contribution to its disrepute betrays a fundamental confusion over the process of perception: " ... I do not know what neurophysiological process there could be whereby neurophysiological processes themselves would seem to themselves - or to other neurophysiological processes - to be located outside the organism altogether."[1]

To give Smythies his due, his strictly neurological efforts are in the main unhampered by his philosophical flights. And even the best of the philosophical neurologists, W. Russell Brain, has not quite escaped the clutches of phenomenal space. In describing the nature of perception he says, "Somehow this coding of spatial information in a spatio-temporal pattern of frequencies is decoded into a static representation of a spatial pattern."[2] Neurologists tend to stress the spatial aspect of neural patterns, and occasionally go so far as to worry because the neural pattern

---

1 Analysis of Perception, p. 74.
2 "Perception and Sense Data", British Journal of the Philosophy of Science, 1960, p. 190.

produced on the cortex when a circle is before the eye is not at all circle-shaped. They seem to be taking seriously Bertrand Russell's famous dictum: "All one ever sees is a part of one's own brain".[1] They would be relieved, one feels, if the pattern were elliptical or even squarish, but no such topological similarity exists. What good would it do for the pattern to be circular? The brain has no eyes to see the pattern. Just as "Hello" can travel through any convolution of telephone wire, so "circle-shape" can assume any spatial characteristics on the cortex. In point of fact, at least in some animals there is an observable degree of spatial mapping in the brain. In the octopus and the frog something like images can be traced from the patterns of stimulation on the first level of the optic tectum. The extent to which there is a similar mapping in the human brain is not known. But the spatiality here is no more important than the spatiality of the signal "spreader" of the perceiving machine described in Chapter 7. The maintenance of some degree of contiguity of signals coming from small areas of the retina is to be expected as a matter of engineering. The axonal branches of single neurones cannot be expected to cover the brain with their meanderings, so the

---

1 The Analysis of Matter, 1927, p. 383.

spatial system that preserves richness of interconnection with a minimum of "wiring" is the natural choice. Parliament would still be a representational body if each Member represented people according to the alphabetical order of their names (MP Smith, representing Aaron - Asquith), but it is much more convenient if each member represents the people of a small geographical area. Neurologists have been sidetracked into hunting for phenomenal space in the brain, a task that philosophers - even phenomenalists - must recognize as absurd. Philosophy, the self-proclaimed guardian of the sciences, has fallen down on the job.

Herbert Feigl, in the very process of arguing for a physicalistic interpretation of mental events, decides to maintain the distinction between physical and phenomenal space.[1] He is led to do this, I would guess, by the credibility and apparent efficacy of at least parts of the Gestalt programme of psychology. Gestalt psychologists have found it very useful to maintain the concept of a phenomenal space, and to describe their findings in spatial terms.

---

1 "The 'Mental' and the 'Physical'", Minnesota Studies, Vol. 11, p. 407-8. Feigl and Sellars proclaim physicalist views that somehow are supposed to accommodate "raw feels" as unanalyzable or irreducible concepts (entities?). And yet these raw feels are not "nomological danglers" apparently. Their rationale or justification for the move eludes me, and I am not alone in this. Avowed disciples of Sellars and Feigl have admitted to me that they cannot explain the steps involved.

Their findings are not to be denied, but their mode of description is pure metaphor. Their phenomenal space or visual field is simply a metaphorical construction in which <u>informational</u> distinctions can be portrayed in <u>spatial</u> terms. Their talk of figure and ground is quite directly translatable into less vague talk about information and signals: "There must exist a degree of heterogeneity between the figure and the ground, each of which tends to form a whole of its own" means that the information received is already about objects out there (on the ground or background), and not about colour patches only. "Protusion of the figure out and away from the ground may also appear" means that signals from the retinal area on which the fixated object casts its image are facilitated or boosted, while the irrelevant peripheral signals are inhibited. "There may exist a halo effect around the figure, which is dependent upon the contour line separating the two" means that facilitation of the signals spills over the edges of the retinal image. As in the case of the frog, the perception of an edge depends on a combination of excitatory and inhibitory signals from both sides of the edge of the image on the retina.[1] What could

---

1 Gestalt conditions from Wever, "Figure and ground in the visual perception of form", <u>American Journal of Psychology</u>, 1927, Vol. 38, p. 194, quoted in George, <u>Cognition</u>, 1962, p. 153.

sound more metaphorical than "protusion" and "halo effect"? One can ask the old embarrassing questions about the figure and ground: what are the dimensions of the figure? how far does it protrude? The Gestalt psychologists have taken introspective reports at face value (which is not unreasonable, but nevertheless a theoretical leap) and their discoveries have thus been expressed in the television-screen mode. But all that need be the case is that the neural signals (in this respect analogous to the microwave signals of television) are related to the <u>retinal</u> image in the way they are; there is no need to suppose them decoded <u>à la</u> television-receiver into another image.

Of course the Gestalt way of talking is fairly safe and very useful until the neurological facts are in, but it should be borne in mind that talk of phenomenal space is metaphorical only, and not at all indispensable to science - even if, when the neurological facts are in, the Gestalt mode of speaking is kept on as a heuristic short-cut.

Feigl raises another point that may seem to put the proposed theory of awareness and experience in difficulty. He points out that having a certain experience and being aware of having it are two different phenomena.[1] Leaving

---

1 Feigl, <u>op. cit.</u>, p. 417. Ryle and Husserl are among the others who make this point.

aside the question of having emotions without being aware of them, which will be discussed later, are there cases of having visual experiences without being aware of them? The analysis depends on how strongly the word "have" is taken. If all that is meant is that I was $aware_2$ of this and that - that signals of certain contents travelled far enough to influence or control behaviour, then of course I have many experiences I am not aware of. A stronger use of "have" would require that later I can report what experience I had - and not just by an inductive inference of the following type: I played the piano yesterday and thus I must have had the experience of raising and lowering my fingers in certain ways, etc. In this sense, to have had an experience, one must remember it. But here again, there is nothing in the proposed analysis to suggest that awareness is required for memory traces to be produced, although it would be expected that signals not strong enough or important enough to cross the awareness line would be unlikely to produce strong memory traces. It is often the case that we do not notice a particular object that passes before our eyes for days on end until the object is removed; then we realize that something has changed. If this is what is meant by having a visual experience without being aware of it, the proposed analysis can accommodate it. And what else could it mean?

It is tempting to say, after introspecting for a while, that it is all very true that there is only a small, central part of the visual field of which we are aware at any moment, and that to describe the whole scene, our eyes, our fixation point, and our "focus of interest" must scan the sensory presentation, but that the parts we are not scanning at any moment persist or remain, as a sort of vague, coloured background. It is this, perhaps, that we experience without awareness. But here introspective description gets into trouble. For as soon as one becomes interested in what is going on outside the beam of the fixation point, one immediately becomes aware (aware$_1$) of the contents of the peripheral signals. And this phenomenon is quite different from the ordinary one. While it is true that one can focus on a spot on the wall and yet direct one's attention to the peripheral signals and come up with reports like "There is something blue and book-sized on the table to my right; it is vague and blurred, and I am not sure it is a book", it cannot be inferred from this that when one is not doing this, one is still experiencing the blue, booklike shape. We are easily deceived into such suppositions by the natural operation of our eyes, which is to make a cursory scanning of the environment whenever it changes and as soon as it changes, and by the operation of our short-term memories, which holds the results of this scanning for reference for a short period of time. In familiar surroundings we do not have to see or pay attention to the objects in their usual places. If anything had been moved or removed we would have noticed, but that does not mean we notice their presence, or even that we had the experience (in any sense) of their presence. We enter a room and we know what objects are in it, because if it is a familiar room we do not notice that anything is missing, and thus it is filled with all the objects we have noticed or put there in the past. If it is an unfamiliar room we automatically scan it, picking out the objects that fill it and catch our attention. The stable continuant behind our scannings is none other than the physical, spatial room itself, reinforced by our memories and hence by our anticipation.

27. Colours.

Still to be explained is the notion of subjective colours and other "secondary qualities". The public, three-dimensional world of every day is filled with coloured objects, but the scientific world of particles has no colours or sounds. It is clear that a complete explanation of colour must be in terms of the particle world - even if parts of the explanation (non-crucial parts) are left in terms of the public, everyday world, unanalyzed (but clearly analyzable) in particle-world terms.[1] Colour, as everyone knows, is not a property of particles or a thin skin of objective physical paint on objects. And yet colour seems eminently spatial ("Everything coloured is extended"), so must there not be some space filled with extended objects covered from end to end with colours? Since colour is a subjective phenomenon, the obvious answer is that the objects of subjective, phenomenal space are the vehicles of colour.

The complexities of the physical theory of colour and its relation to experience have been dealt with at considerable length by Smart,[2] and for once I can avoid the messy, physical, unphilosophical details and simply present Smart's conclusion, with which I agree: " ... all secondary quality concepts concern the classifications of sensory stimuli made by complex neurophysiological mechanisms."[3] One <u>caveat</u> arises however: the only place where classification <u>by quality or physical property</u> goes on is right at the retina, in the light-sensitive cells. (See Chapter 7, on stimulus-checking mechanisms, pp. 151-3.) The relationship between

1 I do not wish to attempt to adjudicate which is the <u>real</u> world; as Quine says, the option is unreal. (<u>Word and Object</u>, p. 265.)

2 <u>Op. cit</u>., Ch. IV, see also Geach, <u>Mental Acts</u>, 1957, pp. 33-8.

3 <u>Op. cit</u>., p. 86.

colours experienced (reports of colours in experience) and wavelengths is not at all, as Smart points out, a simple one-to-one correspondence. A variety of very different combinations of wavelengths can produce the same experienced colour. In neurological terms, a wide variety of wavelength combinations can trigger identical or nearly identical higher-level colour signals. (They are identical, of course, in function - whether or not they are identical in physical properties, paths taken, etc.[1] Smart concludes from this that "There is no reason to expect a close correspondence between these classifications and the way things in fact are in nature."[2] The high-level information that issues in reports about colour is thus not well-described as information about light-waves, since there is no simple relation between these high level signals and any particular light-wave combinations. Similarly, news reports about the

---

1 Smart mentions some of the experimental findings that militate against a simple three-colour theory of vision, but since Smart's book was published new experimental evidence strongly suggests that the first step in colour vision does involve variable retinal signals from three types of differently pigmented cells sensitive to blue, green and yellow. At a higher level these signals are apparently transformed into some sort of positive-negagive, on-off signal system. The simple three-colour theories must be discarded and replaced by a theory of two-stage (at least) complex three-colour systems. (MacNichol, "Three-Pigment Color Vision", Scientific American, December, 1964.

2 Op. cit., p. 86.

number of people on the beaches on various days would be poorly described as being reports about the weather on those days.

It could have been the case that our neural organizations gave us the information quite directly about light-waves, but the world being what it is, this would not have been of as much use to us as the highly complicated system we have (which fact goes a long way toward explaining why we have the system we do - and why fish, for example, have different systems). The great wonder of the human visual system is that it orders the complexity of light-waves so that the multifarious combinations of light-waves that can be reflected by an object under normal conditions are subsumed under one informational heading, so to speak, providing us with a means of recognition of objects on the grounds of light-wave combinations alone.

Smart's contention that "colour concepts do not correspond to anything simple in nature"[1] is, on the other hand, very misleading. If it were true, such everyday operations as identifying apples and daisies and mixing paint could not succeed with the regularity they do. Whereas it is true that a particular experienced colour can

1 Op. cit., p. 84.

be caused by a number of different reflectional situations, and thus experiencing a particular colour is not a sufficient condition in itself for proclaiming a particular simple state of natural affairs to be the case, if a particular simple state of affairs is the case, under normal conditions a particular colour will be seen. Copper solutions always look green to normal observers under normal conditions. The fact that changes in the surrounding environment can alter the experienced colour does not indicate that this particular experience of green is not closely connected to some basic reflective property of copper. The assumption that is unwarranted is that some other green thing, say a leaf, had the same reflective property (at least to a degree) as the copper solution. We might all have been born with visual systems that differentiated all colours except green and red, which were interpreted as one colour. It would then have been tempting on the basis of the facts of visual experience to frame the hypothesis that the reflective properties of a ripe tomato and of a copper solution were the same, or were simply related to a shared fundamental characteristic of their particle structures. But such an hypothesis would be completely unwarranted. Smart points out that there could be creatures with completely different colour differentiations from our own, but he does not mean, I suppose, that these creatures might see copper solution under normal conditions sometimes as one colour and sometimes as another. If we made a non-cupric paint sample that matched (for us) the copper colour, such a creature might find it an entirely different colour, but surely copper solutions would all have the same colour for him.

This makes it very difficult to say just what information about the surfaces of objects is presented to us by colour signals. To oversimplify, a message to the effect that a particular apple is a particular shade of red informs us that its surface has the complex reflectional properties of other red apples, and only perhaps the same reflectional properties as tomatoes or fire engines or stop-lights. Of course it is possible that other red apples are red "for different reasons", but this is highly unlikely. It is not at all certain or even particularly likely that apples are red "for the same reasons" that fire engines are red (unless fire engine paint is chemically similar to apple-skin pigment); this is clear because it is not likely, from the colour-blind person's standpoint, that apples and leaves are red-green for the same reasons.

The fact that colour information is in this way left quite open as to its "actual meaning" does not prevent its being useful in the control of behaviour. It informs us about surface differentiations that happen to make a difference on this planet, not only because other humans recognize the same differentiations and pattern behaviour on them (by following flags of certain colours, agreeing to stop at red lights, dressing boy-babies in blue, and so forth) but because these surface differentiations can be relied upon as indicators of physical difference (showing that fruit is ripe, iron is rusting, snakes are poisonous).

Colour is spatial in that colour information is about surface differences in actual, physical space. The information may lead one astray (if the object viewed is in artificial coloured light, for example). In such cases it is not that the red we _see_ on the white object in red light is not in physical space but in phenomenal space, but that our information is that the object is red, not white.

Sellars speaks of the ultimate homogeneity of sensations, and bases his reliance on "raw feels" on this phenomenon.[1] As I understand him, the fact that objects appear _uniformly_ coloured, continuously coloured, and not made up of coloured particles leads him to propose an ultimate disunity between the objects of science and the objects of perception. But

1 _Science, Perception and Reality_, 1963, pp. 25-37.

consider verbal reports, which exhibit a similar "homogeneity" in that they are limited in the information they can carry. "Smith is wicked" suggests, in a way, that Smith is wicked through and through; at least it does not say that Smith is a collection of particles, all or most of which are wicked. "The nation is prosperous" suggests that prosperity is homogeneous in the nation - or at least it does not explicitly say that it is not. Signals can carry information only as particulated as the thresholds of discrimination allow. A signal is homogeneous in function; it is not a mosaic of different functions - the idea does not make sense. This holds for sensory and non-sensory signals alike.

28. The Sensory Core and the Act of Seeing.

The analysis of perception developed in this chapter, and more particularly the physiology behind it, can help to explicate two epistemological themes that are not directly connected with the idea of phenomenal space.

The first of these involves the notion of perceptual reduction. For years the raison d'être of sense-datum philosophy was the quest for certainty, which was presumed to be achieved if one somehow subtracted the "objectifying" aura in ordinary perception leaving the "sensory core" as remainder.[1] This operation was supposed to isolate the sensory raw material, which was certain, from the interpretational increment of normal perception, which, in its postulation of physical objects existing in physical space, was uncertain. One was left with "round, reddish somewhats" or was "appeared to brownly", and the certainty of one's reports to this effect was unimpeachable. The fallacy of perceptual reduction has been pointed out before,[2] but the physical details of just what is going on when one performs "perceptual reduction" are instructive.

In order to experience anything like a red round patch in the place of an apple one must disorganize one's visual system. The first step is to squint and purposely throw the eyes out of focus. This has the effect of blurring things just enough so that if one did not already know what a particular object in front of him was he might not be able to tell. Without this deliberate interference in the initial retinal stimulation no amount of "clearing the mind" will produce the red-round-patch effect. It helps to close one eye, thus eliminating most of the depth and distance

1 The term "sensory core" is from Roderick Firth, see next note. Cf. Lewis, Mind and the World Order, 1929, and An Analysis of Knowledge and Valuation, 1946.
2 Particularly by Firth, "Sense-Data and the Percept Theory", Mind, 1949.

information that would otherwise tend to give the game away. But even with all this "reduction" one will see, vaguely to be sure, an apple, and not a red, round patch unless one also thinks very hard "red, round patch", or something to that effect, thus overwhelming the action of one's apple-association signals with patch-association signals. One must "set" for patches, not apples. So all that one has accomplished is a distortion of one's sense experience and not a purification.

This is not to be confused with Chisholm's account of the justification reduction, in which one modifies one's perceptual claims to meet doubts and criticisms.[1] When one sees that there is a white rat in the corner, one's only justifications of a claim to this effect are the lesser claims: "Well, I see that there is a white animal in the corner", "I see that there is a white object in the corner", "At least I see what might only be a (physical) light-coloured patch in the corner". This is the only way of discounting one's natural and subconscious (in my sense of sub-aware) interpretation of sensory information. It does not involve a perceptual reduction, but a conceptual reduction, an ex post facto reversal of already completed

1 Perceiving, a Philosophical Study, 1957, p. 62.

automatic judgments. As such it is not a method for achieving greater certainty in reports of experience, but a method for achieving greater probability in reports of reality. The limit is reached, as Chisholm points out, when one discounts all claims to the veracity of one's experience and merely clings to the claim that the experience (of seeing a white rat in the corner, or seeing a white object in the corner, etc.) was had, but at this point, of course, the utility of the claim as a communication of empirical information lapses, except in the case (of probing neurologists and psychologists) where what is of interest is only the experience.[1]

Another way of saying this might seem to be to say, with Smart, that "The sincere reporting of a sensation is one thing and the sensation reported is another thing".[2] This is all right if one limits oneself to the interpretation that is obviously true, viz., utterances are not sensations, but if it is meant to suggest, as perhaps it does for Smart, that there is an operation of checking the sensation (brain process) for characteristics and then, on

1 Cf. Ayer, Problem of Knowledge, 1956, p. 41: "In demanding for empirical statements the safeguard of logical necessity, these philosophers have failed to see that they would thereby rob them of their factual content." See also p. 56.
2 Op. cit., p. 100.

discovering what similarities in characteristics there are, uttering the sincere report, then it is very misleading. It is not the case that we first become aware of something and then (afterwards) we are able to say what we are aware of.

The intervening step that changes what one says from "I see a white rat" to "I see a white object" occurs before the awareness line, and hence changes what we are aware of, in the strict sense that it changes what we are aware that. It might be said that the person who changes his claim is still aware that there is a white rat in the corner - but he might deny it. He might admit that he still believed that the object was a rat (although on less than perfect grounds), but that is different from being aware that the object is a rat. Belief is dispositional while awareness is not. Whenever his belief comes to the fore, so to speak, whenever he expresses it (to himself) he will be aware that the object is a rat. If the person is aware that the object is a rat (a temporal event) and at the same time is saying that the object is just an object as far as he is concerned, he must be speaking "automatically" or reciting. If he is lying, his lying utterance may be closely preceded (or) followed by unuttered signal contents to the effect that the object is a rat, but the signal that crosses the awareness line to produce the lie is not to the effect that the object is a rat but to the lying effect that the object is just an object.

The point I am making is that in Chisholm's justification reduction, what is changed is a description, and this description is changed not by undoing the interpretative work of the brain (which is physically impossible) but by doing <u>more</u> interpretative work. There is no object perceived in any case (save, in ordinary terms, the physical rat), and hence no perceived object (least of all the rat in the corner) is changed by any change in neural processes. There can be a change in description if one distorts one's experience by squinting and so forth, but this is not the change desired.

The second theme which is clarified by this analysis of perception is that experiencing is an active process, and not a mere passive reception of sensory information. This theme takes various forms. In Gestalt psychology it is stressed that there is much that we <u>put into</u> perception. Phenomenology maintains that to experience is to act, and to act is to posit an object of that act. The theme is also implicit in any philosophy that attempts perceptual reduction.

If this idea does not simply mean that perception is a complex physical process, it is not at all clear just what it could mean. The active-passive distinction does not seem to fit the facts very well. Are missile-guiding computer complexes active or passive? The radar antenna revolves, electricity goes here and there, predictions are made and checked. Is this active? Or is it merely a passive complex of passive receptors that mindlessly react in set ways to set conditions?

The distinction in human beings may seem to rest on the notion of volition. Is it not true that we see, in part, what we are looking for, or in other words what we <u>want</u> to see? And can we not turn our eyes on whatever we want, whereas the radar machine is programmed to turn its eyes according to set schedules? But the interpretation in perception is usually added subconsciously and unintentionally (see Chapter 11). A person can "set" for deer or for dogs, but so could a perceiving machine. The claim that to see something is to do something does not yet separate that event from digesting food or shivering.[1]

If this theme is stressed, as I suspect, merely as an alternative or antidote to the notion that we are passively presented with sense-data which we then peruse, I have shown

---

1 Cf. Charles Taylor, "Phenomenology and Linguistic Analysis", in <u>Proceedings of the Aristotelian Society</u>, 1959, Supplementary Vol., p. 95.

how this notion can be opposed without mention of action or volition. Whatever else of value may be suggested by the theme of action will be taken up in Chapter 11.

29. Sounds, Smells, and Itches.

The results of the analysis of visual perception can now be applied to questions about the other senses quite automatically. The status of sounds, smells, itches, and pains can be seen to be parallel to the status of images, with pains differing from the others in being more directly and strongly connected to responsive behaviour. The drumming on the tympanum is the last level of sound reproduction. No inner nose smells reproduced odours, no inner foot itches. The case of itches is revealingly parallel to the case of colour vision. Two itches that seem just the same to a person (which only means: they are, as pains, equally intense and he uses the same words to express the signal crossing the awareness line) may have quite different causes. One may be triggered by tweed on skin, the other by some unpleasant chemical, and yet the same signal, functionally, is produced. This is no different from the case to two quite different objects both looking red "for different reasons".

The theoretically safest idiom for reporting sounds, smells, and itches would be "I hear that Jones, who is hoarse, is saying 'Hello'", "I smell that there is Chanel No. 5 near my nose", and "I feel that something is making my foot itch". Particularly in the case of "noises", which approach the complexity of visual "images", it would be best to give descriptions in the "hear that ... " form. Audio information is highly interpreted; it can be analysed (not perceptually reduced to its parts) only by expert musicians and the like. We do not hear all the overtones that go to make up a familiar voice, any more than we see all the angles and curves that make up the retinal image of a familiar face. Only the acoustical engineer knows by what overtones he can tell the brass from the oboe.

All these idioms can be put in the "I seem to ... " idiom except pains. Pains are already in the idiom, in that to say one has a pain is not to say one has an injury. "I seem to hear a voice" is used for disallowing all mistakes and illusions in the path from ear to awareness line. Pain is different. One actually has a pain if the signal occurs, veridical or not - whether or not there is something injurious causing the signal. It could have been the case that instead of using our ordinary pain idioms we always said "I seem to have an injury" or simply "I have an injury". The latter could be shown to be false occasionally, and then one would fall back on the former. Injuries are to pains as white rats are to visual signals crossing the awareness line, and as sound waves are to auditory signals crossing the awareness line. "I seem to have a pain in my foot", although quite a normal sounding sentence, is equivalent to "I seem to seem to have an injury", and parallel to "I seem to seem to see that there is a white rat in the corner". The added "I seem" does not disallow any further possibilities for error and is therefore a totally useless appendage, if not meaningless.

There is good reason why ordinary language should develop in this way, picking out pain-reports for special treatment. Pain signals are more often hallucinatory and more often tell us little about what injury we have. We hear trumpets and see trees but, lacking the medical knowledge until recently, we do not feel ulcers or sprains or arthritic swelling. Pain signals are of prominence because of the strong effect they have on behaviour. It is only to be expected that the idiom that arose placed pains at a different logical level, and thus it is not to be wondered at that pains have been considered perfectly real entities existing in feet, hands, and heads. Distinguishing pains from injuries was to be expected however since location of pain and location and description of injury have always been separate.

The man in the street is not advised to change his ways and adopt the "that" idioms, unless he wishes to theorize about experience. Talk about images, odours, and tingles is informative in spite of its fundamental ontological confusion. To some extent the man in the street does use the "that" idiom, when he says that he sees that ..., or imagines that ..., or hears that ..., but there is no reason to applaud this usage, since the alternate usage leads to no more <u>practical</u> difficulties.

CHAPTER 10.

Reasoning.

30. Is Reasoning Mental Activity?

Psychologism, the view that the elements of logic are ideas, and the laws of logic empirical laws about how we think, has been well and roundly demolished by a series of writers, starting with Frege.[1] No one now holds the view, but perhaps misinterpretation of the arguments against psychologism has led to an equally strange view - never quite expressed, but seemingly implicit in some writers - that logical operations need not at all be realized in temporal operations in order to work (for one to deduce, infer, syllogize or reason).

When computers are made to perform logical operations, the abstract, timeless transformations and connections are realized in physical, temporal operations, and the production of results or conclusions takes time and energy. Is there a human counterpart for these physical operations?

Ryle says, in his ridiculing tone of voice:

> We hear stories of people doing such things as judging, abstracting, subsuming, deducing, inducing, predicating and so forth, as if these were recordable operations actually executed by

---

1 See especially his Foundations of Arithmetic, pp. 36-8, and Husserl's Logische Untersuchungen.

> particular people at particular stages of their ponderings. And since we do not witness other people in the act of doing these things, or even catch ourselves in the act of doing them, we feel driven to allow that these acts are very subterranean happenings, the occurrences of which are found out only by the inferences and divinations of expert epistemologists. These experts seem to tell us that we do these things somewhat as anatomists tell us of the digestive and cerebral processes that go on inside us without our knowledge.[1]

In a way, Ryle has hit the nail on the head. The logical operations that occur when we are reasoning are cerebral processes, of which we are not aware. We are aware of the contents of certain signals; we say to ourselves "Therefore I conclude that p", and "if p, then q". Clearly, just saying these statements is not reasoning or performing logical operations. Ryle argues that such quasi-logical verbs as "conclude" and "deduce" are for use only in the presentation of results already arrived at. The only time one concludes that Smith is the murderer is when one says or writes or thinks to oneself: "and so I conclude from this evidence that Smith is the murderer". This use of "conclude", Ryle says, should not be seen to indicate the existence of a concluding-operation occurring in the mind or brain. Is he suggesting that there are logical operations behind the statement, but they are not

1 Concept of Mind, p. 285.

concludings or deducings, or that there are no temporal, logical operations involved in human reasoning at all? The latter alternative is just not tenable.

Miss Anscombe, in discussing the rôle of the practical syllogism in giving one's reasons, concludes: "But if Aristotle's account [of the practical syllogism] were supposed to describe actual mental processes, it would in general be quite absurd. The interest of the account is that it describes an order which is there whenever actions are done with intentions ... ".[1] An order which is where? Inhering in or ordering what? The usual interpretation of her remark is positively mystical: it is just an order that there is, and it comes out in asking for reasons and giving reasons, but it is not an order of temporal progression of occurring logical operations, and there are no mental or cerebral processes that are thus ordered. Another interpretation, not overwhelmingly suggested by Miss Anscombe's writing, is that running through a syllogism in one's head is not by itself reasoning, but that the syllogism exhibits, perhaps only partially or in an idealized way, the order of logical operations realized in some mental or cerebral processes (perhaps subconscious ones) when one reasons.

---

1 Intention, 1957, p. 80.

The point that survives, when Ryle's and Anscombe's arguments are interpreted with sympathy, is that running through syllogisms or saying "therefore p" in one's head is not reasoning and is not the human counterpart of the temporal realizations of logical operations in computers; it is at best an expression that somehow exhibits the order of some cerebral functions that have occurred, but of which one is not aware. Their warning should be, not that searching for logical operations in neural function is hopeless, but that one should not look for a collection of neural concluders, premise staters, and deducing mechanisms.

Frege's point was not that there is no such "mental" activity as reasoning, but that logic is not psychology, any more than arithmetic is the examination of calculating machines. The notion of a "physiology of logic" is not ridiculous; the scientist who takes apart a computer to see how it works would be in effect studying just that - computer "physiology".[1]

---

1 Karl Lashley, for one, recognizes this. See "The problem of serial order in behavior", in Jeffress (ed.), <u>Cerebral Mechanisms in Behavior</u>, <u>The Hixon Symposium</u>, 1951.

31. Operations with Information.

The physiology of logical operations with information in the human brain is a science not yet born, and even when it is, it will be enormously complicated, dealing as it will with an apparatus much larger, more versatile, and more changeable than any computer. But there are some points that can be made of a rather *meta*-physiological sort about storage of information and how it must affect operations.

Storage of information in the brain cannot be simply a matter of filing away some jot of information that can then be drawn from the files, in a manner of speaking, and brought across the awareness line. There is nothing to the information carried by neural signals but function, and storing information must be making changes in the functional potential of some part of the brain, so that later signals are operated with in new ways. If one must speak of the vehicle or language in which the information is carried, then the language can only be the functions of the neurones involved. There is no coded message to be recorded and stored, like a book in a library, and then indexed for retrieval.

Storage, moreover, is not to be separated from retrieval. Information supposedly stored (with storage defined in terms of some physiological changes) which turned out to be impossible to retrieve for use, is not information stored, but information lost. Just as the computer is in state A if it does what it is supposed to do when in state A, and is not in state A otherwise, so the neural computer has the information A if it functions as if it had the information A, and not otherwise.

Simply on intuitive grounds it is clear that a functional change amounting to information storage cannot be just the establishment of the capacity to say that A or "I know that A" when suitably stimulated. The information that fire hurts would not be said to be stored by a brain if all the person could do was say "fire hurts" - if he continued to get burned, and behaved inappropriately when in the presence of fire. The brain's storehouse is not a collection of tape recordings for the use of the speech centre, but an integral part of the controlling mechanisms for the whole body. In ordinary parlance we would say that a person must understand what he knows; for this to be true, storage and the wide range of human controls must be intimately bound together.

It may in fact be the case that there are storage areas of the brain anatomically distinct from the operational or control parts of the brain, but even then the actual storage of information - in the sense: storage that can make a difference to behaviour, knowledge, reasoning, and so forth - must involve the establishment of neural organizations to retrieve, utilize, and "understand" the stored information. Understanding in this metaphorical sense can only mean that the stored information is a certain functional potential integrated with the functions of the controlling areas.

To oversimplify, when the information that fire hurts is stored in the brain, this cannot mean simply that one can say (if one has learned to speak) "fire hurts", but that future jots of information about the presence of fire, for example, trigger motor activity designed to allow the person to avoid the fire. Or, if the jot of information "cigarette embers are fire" arrives, the organization set up by the previous information that fire hurts produces the new jot of information that cigarette embers hurt, and a new organizational change. Thus the addition of information B to the stored information A produces a third bit of information C. If the arrival of information B does not have this effect, then information A was not stored in the first place. These capacities must exist to some extent in dumb animals as well as in man, with the difference only that no new verbal capacities are produced in the animal, since it has no verbal capacities, no verbal behaviour.

In humans, information stored need not come to the fore, need not be brought across the awareness line in the form of messages, to be used. Only the resulting signal of a production of new information may cross the awareness line, or it may in turn trigger the production of still further signals, and these may or may not cross the awareness line. If I have stored the information that tomorrow is Friday, and I see on the calendar that on Friday we are dining out, I can say almost immediately that tomorrow we are dining out, without running through the argument in my head. But I cannot do this _just_ as soon as I see the calendar; the information from the calendar must be transmitted into the storage-operational areas which produce the "conclusion" that I can say. The signal with the content "tomorrow we are dining out" may instead go into another organization outside the awareness line, with the result that the signal that crosses the awareness line produces the utterance, for example, "do I have a clean shirt for tomorrow?". Whatever one wishes to call these subconscious productions of new information, their operation is essentially logical, and they must occur if behaviour is not a divine and impenetrable mystery, which it is not.[1]

---

1 I may be "reminded" that tomorrow we are dining out by something irrelevant: a faint smell of cedar, say. How this stimulus operates to bring my stored information

Both knowledge that and knowledge how, to use Ryle's terms, must be stored in such a way that future information can be operated on to produce new information, new knowledge, and new behaviour controls.[1] When one knows how to add or subtract there must be neural organizations that will normally produce the right answers when the stimulation, the information, is fed in. One might say that all information is stored as knowledge how, since, for example, if I have stored the information that tomorrow is Friday, I have the capacity or ability to transform new information about Friday into information about tomorrow. But this description should not be taken to mean any more than that I have in my brain neural organizations which, so to speak, do the transforming for me. Knowing that tomorrow is Friday is not very much like having a talent or knack, but the difference is not a radical one. It might be supposed that knowledge <u>that</u> is absolute in a way that knowledge <u>how</u> is not: with knowledge that, you either know it or you do not. But this is not quite true. There is relatively little of

---

about dining tomorrow to the fore is not apparently "logical" in any way, but the logical associations must exist if I can go on to say (to myself, perhaps) "tomorrow we are not dining at home".

1 The concept of knowing is not without difficulties. Prior to Chapter 12, I shall attempt to use the word in non-problematic contexts only.

interest that I can do with the knowledge that table salt is sodium chloride, while the same information can be of immense use to a trained chemist.

Ordinary everyday information has approximately the same functional potential for everyone, and hence the statement, said of an adult, that he either knows that p or he does not, fits the facts quite well. But with young children, whose understanding is partial, the alternative is not so ironclad. Is it so clear that the four-year-old who says "Daddy is a doctor" either knows or does not know that his father is a doctor? Does he know what a doctor is, or what a father is, for that matter?

32. Logical Necessity.

How do the right organizations almost always happen to set themselves up in the brain? How does the influx of information work to produce new information according to the laws of logic? The answer for the computer is simple: it is designed that way, as an embodiment of logical operations.[1] No such answer will work for the brain.

1 In the computer, both the atomic activities of the individual modules and the activities composed of these atomic activities could be construed as "logical operations". But even if the functions of individual neurones are seen as "logical" (since they might be described as signalling the alternation of different conjunctions of stimulus conditions), the composite activities are certainly not all logical in any sense. Recognition, for example, is not a logical operation or activity.

It is perhaps tempting to suppose that since, as logicians say, logical transformations depend on the defined capacities of the logical operators, such as "V", " ", and "(x)(Fx)", the ability to think logically or reason depends on knowing how to use the ordinary logical operators, "or", "if-then", "all", etc., and the ordinary class-logical expressions, "is a member of", "belongs to", "is in", etc. But this will not do at all. Animals are capable of combining information logically (at least, they combine information to control usually successful new behaviour), and children can "put two and two together" to arrive at new information long before they have mastered the ins and outs of the logical vocabulary. Furthermore, there is no set list of strictly logical operators in ordinary language, and those that there are are often left out of ordinary conversations (Consider the use of "since", "because", and "so" in ordinary discourse at the expense of the standard logical operators, and the unsignalled logical distinction between "Lions like red meat" and "Lions are getting scarce",[1] a distinction that is not lost on little children, and yet not noticed by them either.) Children deduce and conclude and weigh alternatives before they learn

1 Cf. Quine, Word and Object, Sections 28-9.

the use of "deduce", "therefore", "or", and certainly before they have mastery of the if-then operator.

I have explained earlier how "non-logical" association organizations can be set up by the force of intra-cerebral evolution - learning from experience. "Logical" associations can be no different in means of establishment. Logical truths are, after all, general truths with myriads of particular instances available in experience, so there is no reason why logical associations should not be set up "inductively", like non-logical associations.[1]

The fact that some of the information we store and use is necessarily true does not alter its function. One can discover empirically at first that triangles - at least all the triangles one has ever examined - are rigid figures, and thus use this information in neural organizations controlling various activities. I should imagine that we all learn by induction at first that when you put a box in a

---

1 Many philosophers reject the notion of a strict distinction between logical or necessary truth and merely empirical or synthetic truth, and although I agree with them, the point requires considerable argument. So for the purposes of this argument, I shall consider the stronger view (in terms of what must be explained away) that there is a strict difference. That is, my argument here is independent of the competing dogmas about necessity. (See Quine, From a Logical Point of View, especially "Two Dogmas of Empiricism", reprinted from Phil. Review, 1951; Feigl on nomological nets; and Grice and Strawson, "In Defense of a Dogma", Phil. Review, 1956, pp. 141-58.)

box in a box, the first box is in the third, that when Teddybear is mine and Teddybear is a toy, Teddybear is one of my toys, that it is quicker, <u>ceteris paribus</u>, to walk to school in a straight line. One learns these facts, and their more generalized forms, only as empirical facts on a par with such facts as the fact that fire hurts. The "effect" of the learning may be stored without ever becoming part of a conscious or deliberate theory. The leap from a collection of particular associations to generalizations is well within the capacities of neural nets (see Chapters 4 and 5), and is revealed in more or less atomic form in such phenomena as Fuchsian completions (see p. 206). Such "jumping to generalizations", once performed by neural organizations, would be well-rewarded in experience and hence reinforced and encouraged as a neural move. The utility to behaviour control of stored necessary associations does not rest in their necessity, but in that they invariably, empirically work. The man in the street need not know that some of the things he knows or has stored away for use are necessarily true, and in fact he often will not have any ideas on the subject or even realize that he knows these things.

It is even possible, though highly unlikely, that some basis of logical operation, established in the primeval days of evolution, is an inherited part of the infant brain's programming.[1] The slave boy of the _Meno_ may in fact have some "inherent knowledge" of geometry, but it is more likely that he learned what he did "subconsciously" in the course of his previous experience, and in the course of being prodded by Socrates. The source of his knowledge - and his knowledge is incidentally quite a lot like a talent or knack in this case - has no relevance, of course, to its being necessary or synthetic.

It should not be forgotten that human rationality is far from perfect, which would not be expected if each man were equipped with a neural computer for churning out logical computations. Unlike most actual computers, human beings reach incorrect conclusions not just as a result of physical breakdowns, but as a result of faulty design, or at least so it would seem when one considers the persistence and vehemence with which people can cling to the conclusions of their own demonstrable _non sequiturs_. Feedback and recalcitrant experience could tend to correct simple breakdowns immediately, and in the long run even faulty design - but only in the long, exasperating run. The frequent

---

1 Cf. Turing, "Computing Machines and Intelligence", _Mind_, 1950, reprinted in Anderson, _Minds and Machines_, p. 26.

truth of the conclusions of non sequiturs must dissipate the force of experience in correcting faulty design. Thus although it is known that McCulloch-Pitts neurones in proper organization can serve as a digital computer or Turing machine, nothing so simple as a well-designed digital computer of such neurones is to be expected in the human brain.

Does this transformation of new information on the basis of stored information, a process open to men and animals alike, deserve to be called reasoning? If it is held that reasoning must involve the use of language, then what animals do is not reasoning, but then neither is a great deal of patently rational human activity that does not depend on use of language. If it is held that reasoning must involve the "apprehension" of connections as necessary, then human beings seldom reason, unless they are logicians or philosophers of certain schools. If reasoning is held to be the transforming of information by operations that happen to be necessarily appropriate then animals can reason along with men. They cannot of course give their reasons or produce a proposition as the outcome of their reasoning, but that is just because they cannot behave verbally.

Animal behaviour is generally successful, and it is successful because it is appropriate to the information that determines it. The appropriateness and hence the survival of neural transformations depends on the way things are, and part of the way things are is the way things logically or necessarily are. The neural operations occurring in an animal - not counting those operations on their way to extinction - are logically appropriate. The operations do not produce logical truths to be uttered, but they produce appropriate behaviour, which in human beings can include the uttering of truths, logical or not. The statement "tomorrow we are dining out" is not a logical truth, but my verbal behaviour of saying "tomorrow we are dining out" is produced by logically appropriate operations on the input information.

There is no need for elaborate experiments to show that dumb animals operate logically with information. The everyday fact that a dog which has been burnt by fire a few times avoids fire shows this. Consider the behaviour of certain low-nesting birds that feign a broken wing when a predator appears, in order to lead it away from the nest where the unprotected young are. It is possible and even likely that the bird's behaviour is instinctive and inherited, and thus not an example of producing successful new behaviour on the basis of incoming information (the sight of the predator acting as a simple trigger for the instinctive behaviour). But how is the predator's behaviour to be explained? Is it, too, instinctive? Have ages of evolution impressed on its brain the behaviour of pursuing limping birds? It is unlikely, considering the infrequency of occurrence and the regular fruitlessness of the behaviour. The bird's trick would not work if the fox or dog could not act rationally, unless of course the fox's or dog's behaviour is pure reflex, and the entire performance is a stately ritual dance instinctively performed by hungry predator and alarmed bird, with no benefit accruing to the predator.

One might wish to adorn this behaviour with the postulated "mental process": "I like to eat birds, therefore I like to eat limping birds; I cannot catch flying birds, but this bird is not flying; it is limping; therefore ... ", but that just confuses the issue. Saying the words aloud or internally is not reasoning, but a product of reasoning; chasing the bird would be a product of reasoning in a man, so why not in a fox? Pulling a baby away from a fire could be a product of reasoning in either a man or a dog. In neither case would logical truths be uttered or contemplated, and the resulting behaviour would not be logically true or false, or true or false, but appropriate.

33. Pondering.

But perhaps there is another type of activity which we call reasoning, which men can do but animals cannot. The operations described so far are in some sense swift and automatic, but men sometimes ponder over a problem for some time until they find the solution or give it up. Is this type of behaviour radically different from what animals can do?

The possibility that this type of behaviour may depend on linguistic capacity can be fairly well excluded from the start. In experiments in animal problem solving, higher animals often at least seem to be "sitting and thinking about the problem" before acting. Jonathan Bennett, in Rationality, holds that the most likely explanation of this behaviour is that the animals are engaged in "mental trial-and-error".[1] He makes the point that there is no radical difference between trial-and-error imagining and trial-and-error behaving. The difference is in convenience and speed. Human beings engaged in the same activity: "Sometimes we do this with words, and sometimes by a kind of imaginative and experimental picturing of the outcome of various possible courses of action."[2]

---

1 Rationality, 1964, pp. 114-9.
2 Ibid., p. 117.

This is, however, far from being a complete explanation, for how is it that the outcome of a course of action follows in the imagination once we have imagined the course of action? I can imagine, especially in dreams, very bizarre outcomes of imaginary actions. I can imagine picking up a teacup and moulding it in a twinkling into a live rabbit. The fact that in problem solving these unusual outcomes are excluded from the imagination - even though I may imagine an outcome which is not right - needs explanation.

Imagining a course of action does not include the outcome automatically if there is anything new or puzzling involved. Although I might well imagine myself turning the key in my front door and having it open all as one step, since it has happened in my experience so many times, I cannot imagine the outcome of picking up a bow and arrow and trying to use it, without considering it in steps and figuring out, somehow, the intermediate stages and the likely outcome. The mere fact that imagination is neither a direct transcription of something earlier experienced (in which case it might be stored and then re-run like a film) nor a completely disconnected sequence of "imagery" must mean that it proceeds in a regulated way, guided by stored information on experience <u>in general</u>. Pondering, as a form of imagining, is either rational or irrational; the activity itself can be neither reasoning nor a prerequisite of rationality, any more than the activity of playing chess could be. Since animals can be supposed to be capable of imagining (without being aware$_1$ of the content of their imagining), they could imagine in a regulated way just as they behave in a regulated way. The intuitive description then, that the animal "sits and thinks out" the solution to problems is not clearly fanciful.

Nevertheless, men are better at this activity - even without language - than other animals are. The cause of this is no doubt part of evolutionary history. Man cannot run as fast, jump as far, or defend himself as well as other animals, but he has opposable thumbs, so he quite naturally and inevitably develops a wider range of controlled behaviour to go along with his sophisticated physical resources. This in turn allows his imagining to filter through more organizations. Man is thus much better at pondering than horses or dogs are, but without language (as in young children) he is not remarkably better at pondering than other primates with opposable thumbs. The more complex behaviour is, the greater the brain capacity becomes, which in turn breeds more complexity. Man's natural vulnerability is balanced by neural organizations which apparently will adapt on less repetition of stimulation, allowing faster learning and the storage and use of more recondite information.

34. Language and Reasoning.

The ability to use language greatly increases the ability to reason, mainly because it enlarges the world of useful stimulation, or in other words, enlarges the world. Since we are speaking creatures, words can stimulate signals having special efficacy in the circuits that determine behaviour. Words as words understood, not as mere sounds, are the added elements of experience. We can be saddened (in terms of behaviour: reduced to tears, lump in the throat, etc.) by lugubrious music, a desolate vista or a face recalled. In these cases words play no rôle. Or, we can be saddened if someone says to us "Smith just died". This will not sadden us because the words have a lugubrious sound. And it will not sadden us unless we understand English and know Smith. We may start if someone springs out and says "boo!", or if someone says, in an even tone of voice, "you're looking however what today", or if someone says, again in an even tone of voice, "Good morning. I just murdered my uncle". The first does not depend on verbal understanding, the second depends on an intermediate degree of verbal conditioning, and the third depends strictly on verbal understanding. A dog might start at the first, a computer programmed to translate into Russian might "start" at the second, but it would not start at the third.

Programming the brain for language cannot be simply programming it to respond verbally to verbal stimuli (as a translating computer might). The point is obvious, not only because the human brain, unlike the translating computer, must have its language programmed into it through inculcation "in the field" where non-verbal stimuli are the foundation to build on, but also because if language were an end in itself and not a means to better manipulation of the environment and thereby to survival, its arrival on the evolutionary front would be inexplicable. From the mechanistic point of view, the programming process must involve convergence of signals from both verbal and non-verbal stimuli, and neural activity beyond the point of convergence must eventually be such that either verbal or non-verbal stimuli alone tend to excite the established associated areas of the other. But the heard words need not conjure up the image of the object referred to, and seeing an object need not bring its name to our lips. What "thoughts" come into awareness is not to the point; what happens on the far side of the awareness line, in the central control areas, is to the point. The neural activity on hearing "dog" must share certain hardly namable features with the neural activity on seeing a dog, but only necessarily at the high level of controls, where information from all sources - linguistic, visual, tactile - is stored and used. Imagination and awareness are not essential to control, although they are closely conjoined with it.

In the long run, behaviour, verbal or not, can be appropriate to verbal stimulation only in that it is mediately appropriate to non-verbal stimulation. The foundation of language programming, and the guarantee of its survival, is non-verbal stimulation, and the raison d'être of verbal behaviour is to contribute to the control of non-verbal behaviour in others. This point is easily obscured when one considers such sophisticated verbal behaviour as expounding metaphysics or writing poetry, but it has its counterpart expression at this level in the empiricist-positivist creed that talk of objects with no empirical foundation, i.e., with no non-verbal sensory stimulation for foundation, is nonsense.

The empiricist dogma, however, is normative in a way I am avoiding. In saying that talk which is non-empirical is nonsense, the empiricist can only mean that it is fatuous or harmfully misleading. "Nonsense" is not a technical term but a pejorative. To be fatuous, for the empiricist, is to be in no way appropriate to non-verbal stimulation and hence to be inappropriate to the direction of any human non-verbal activity. So the empiricist sets out a programme for abolishing non-empirical talk. My point, on the other hand, is that in fact the reinforcement and survival of any verbal behaviour depend on its empirical link and on its appropriateness. The brain can be fooled, at least for a while, into adopting inappropriate verbal behaviour, but the intra-cerebral evolutionary process must work to extinguish this behaviour in the long run - even if it is communally adopted. As verbal behaviour becomes more and more indirect and sophisticated, however, it is affected less and less by the punishments and rewards of appropriateness, and thus it is certainly possible for evolution to lose its grip so that mass linguistic inappropriateness at a high or theoretical level might prevail. Such an Orwellian development would indeed be a catastrophic bewitchment of man's intelligence by language, but the day has not yet come, and the language of everyday coping will act as a good standard-setter into the indefinite future. The empiricist sets out to aid evolution by making clear the inappropriateness of certain ways of talking. He thus is playing Lenin to Darwin's

Marx, abetting the natural process by making its results a goal.

Information that enters the brain via verbal stimulation contributes to reasoning, prompts pondering, imagining, fear and trembling, and contributes in other nameless ways to behaviour control. There are differences of course in the responses to verbal and non-verbal stimuli just as there are differences in the responses to tactile and visual stimuli, or in responses to seeing something dangerous from afar and seeing the same thing close at hand. Verbal stimuli also differ from non-verbal stimuli in presenting the brain with different kinds of information. Verbal stimuli can be requests, orders, vows, questions, and so forth, and although these can be construed as giving the brain information <u>that</u> "Smith wants me to ... ", "Smith wants to know ... ", "Smith's verbal behaviour indicates that ... ", these jots of information are not simple substitutes for non-verbal information. This talk of <u>kinds</u> of information is not very satisfactory, but I do not intend to make anything hang on it. What is clear is that stimulation can be viewed as acquiring its character (request, order, report, etc.) from the way it usually influences behaviour, and since the behaviour in language "games" is structured to be controlled in rather artificial ways by verbal stimulation, the information imparted by this stimulation often has a character quite unlike that of any non-verbal stimulation.

Language creates an environment of stimulation that is more manageable than the environment of non-verbal stimulation, which is its foundation. Words can be used to bring the brain's reasoning facilities into play in hypothetical situations. Where for an animal or pre-linguistic man a problem had to be right at hand before pondering could begin, language enables reasoning to occur at a safe distance or ahead of time, so that appropriate behaviour control can be established in advance.

A perhaps more important gift of language to reasoning is the capacity for information convergence that cannot occur in non-verbal stimulation. One cannot see a chiliagon <u>as</u> a chiliagon because of the limitations of the eye and the neural interpretation of visual signals, but the verbal hybrid "thousand-sided figure" can be used to produce information for reasoning. Thus we can "conceive" of chiliagons, n-dimensional spaces, and Average Taxpayers, even though we cannot see and hence cannot imagine them.

It can be misleading to talk of thinking just with words or concepts as opposed to thinking with images since this suggests that in one case words and in the other case images are manipulated. But the only difference between a "concept" and an "image" is one of source. It is not that one is spatial and the other is not. The "conception" of a chiliagon and the "image" of a pentagon are both to be construed in terms of neural functions in such a way that the spatiality of these functions is beside the point. The difference lies in the different capacities of the visual part of the brain and the part of the brain that has verbal input. The visual part cannot transmit "chiliagon" for much the same reason it cannot transmit "bacon-flavour" or "whistle"; it has no facilities for these inputs.

If someone is talking to me about triangles it may help me to understand what he is saying if I imagine a triangle (by reflexive stimulation of some high level visual signal with that content), but to imagine a triangle is to back-track; it is to feed into the neural organizations information that could just as well have been provided by the initial verbal stimulation, had I been more completely programmed for the use of the word "triangle". If someone begins a tale "yesterday my brother was attacked by a dog ... ", the well-programmed listener can understand what follows without having to "conjure up an image" of a dog, or of biting, or brother, or yesterday (and what could an image of brother be like?); his understanding is a matter of using the verbal stimulation as he would use non-verbal stimulation. Thinking just with words is using verbal stimulation straight, instead of first "translating" it as best one can into visual stimulation by reflexive triggering of high level visual signals, but the information in either case has approximately the same content and goes through the same cerebral mill. Similarly, thinking in French differs from thinking in English and then translating (or hearing in French and then translating into English before thinking) but the signals that occur, for all their content, are not in French or English, are not verbal or pictorial. The brain is not, in its operations, a manipulator of words.

With language must come a basic functional change in the neural organizations that store information. Not only must they transform arriving information into appropriate controls of non-verbal behaviour, but into appropriate controls of verbal behaviour as well. It becomes important to be able to express the results of information transformations, for example, and so the organizations must grow new arms, so to speak, which the organizations in non-linguistic animals do not have. The organization in a dog that stores the information that fire hurts differs from the corresponding organization in the language-using man in being able to use new information in the control only of appropriate non-verbal behaviour. The human organization must be able to control not only appropriate non-verbal behaviour but also the utterance of such expressions as "fire hurts" and (if the presence of fire is announced by new information) "this will hurt".

35. Animal Reasoning.

Jonathan Bennett argues that rationality is dependent on the ability to use language.[1] He arrives at this position by arguing first that "the idea of rationality is that of the ability, given certain present and particular data, to unite or relate them with other data in certain appropriate ways." (p. 85.) This certainly is very much in agreement with what I have been saying about reasoning, although he arrives at his conclusion in quite a different way. He then argues from this that the ability to relate data in this way depends on the ability to make dated judgments and universal judgments. But, he goes on, "only linguistic behaviour can be appropriate or inappropriate to that which is not both particular and present." (p. 87.)

Bennett says: "The only way in which behaviour can either fit or fail to fit a general fact is for it to say something which means that some general state of affairs

1 Rationality, see esp. p. 86 ff.

obtains; and the only way in which behaviour can fit or fail to fit a fact about the past is for it to say something which means that something was the case in the past." (p. 87.) But this is surely false. The very success of behaviour in coping with what is present and particular depends on past experiences which are particular, and have been generalized in neural organizations for behaviour control. If _depending on_ facts is not _fitting_ facts, then Bennett's notion of _fitting facts_ is strict indeed - more like _expressing facts_. If a dog keeps burning its nose on a fire, does its behaviour not fail to fit facts about the past - its previous experiences with fire? And if it avoids fire does this behaviour not fit the facts of the past?

Bennett makes a lesser point, that what can be _manifested_ in non-linguistic behaviour can only be belief in something particular and present. This will stand if one does not consider repeated manifestations of the same belief about what is particular and present as a manifestation of belief in something general. Thus it is true that a dog cannot manifest a belief that toadstools in general are poisonous by refusing to eat one toadstool on one occasion, but can he manifest this belief by repeatedly refusing to eat toadstools? Bennett makes external behavioural manifestation the last word in evidence, but I have been holding that manifestations of internal neural operation should also be considered. The discovery of a neural organization established by past experiences with toadstools that inevitably triggered avoidance of toadstools would surely be evidence that the animal's behaviour was appropriate to what is general and past, whether or not one accepts that repetitions of behaviour manifest belief about what is general and past.

And what, after all, is manifested by the behaviour of saying something about the past or about something general? A parrot can say "John's gone to London", but the verbal expression ensures nothing. One would only be convinced that the parrot's behaviour was appropriate to this past event, if the parrot could do many other things besides saying "John's gone to London". (See Chapter 12.)

Bennett concedes a bit more. It is true, he says, that an animal can manifest belief about what is general, but only if it manifests at the same time belief about what is past; a dog can retrieve a bone it has previously buried, and "the dog's retrieval of the bone manifests, apart from its desire for a bone, two things at once: its belief that it buried the bone in that place, and its belief that in general buried bones stay put until retrieved by the burier." (p. 88.) Bennett stresses the inability of the dog to manifest either of these two beliefs separately, and then concludes from this that we cannot with justice say that the animal's behaviour is appropriate to one or the other.

This is a rather forced conclusion, and Bennett seems to rest it on the notion that without separation of the two kinds of knowledge there is no possibility of rational behaviour. Earlier he says that rationality is the ability to relate present and particular data with other data in certain appropriate ways, to create a "multum in parvo". "For there to be a 'multum' one must at one time intellectually possess more particular data than are present to one at the time, and for it to be 'in parvo' one must have rules or universal statements under which the particular data of which one is possessed can be subsumed." (p. 85.) But "intellectually" possessing non-present data (by which he must mean data not at that moment arriving from the sense organs) is, I should guess, having stored information. It surely is not thinking of all these data or being aware of them all at once. And having stored information is "having rules or universal statements", in that it is having an organization established by past data for the appropriate processing of presently arriving data. There is no other way in which information could be stored in the brain; the brain has no written or spoken language in which records might be kept. Information is not stored or "intellectually possessed" unless it is "subsumed" and "related" in just the way Bennett is requiring. Without this phenomenon of past data changing the functional potential of the neural organizations which will handle future data, neither language-user nor dumb animal could act rationally. One does not report to oneself specific data from the past, and then report to oneself a universal statement of subsumption, and then, as if by magic, reason with incoming information. The function in animals (which is also a function in human beings) of combining the past particulars into generals which control present particular behaviour, far from being a disqualification of rationality, is a prerequisite. What is crucial is not the separation of general and particular that Bennett requires, but their conjunction in established neural structures. The additional capacity for reporting specific data from the past only becomes useful with the advent of language, and is seldom itself used in reasoning. Reports of what is past and particular are useful finally only in that it is useful to communicate to other language users what is past and particular so that they can "subsume" these particulars in their own organizations. The value of parables is in instilling precepts.

Bennett holds that non-linguistic animals can be <u>intelligent</u>, but not <u>rational</u>. (p. 43.) To be rational one must be able, he says, to give one's reasons, which involves mention of that which is general or that which is past. But the importance of giving reasons is only that it can contribute to the reasoning ability or raw material of other communicating creatures. Bennett's criteria for rationality are these: (1) that an animal can modify behaviour on the basis of stored information, and (2) that an animal can communicate its reasons. Aside from the fact that these criteria do not seem to determine the notion of rationality from which he starts (the ability, given certain present and particular data, to unite or relate them with other data in certain appropriate ways), they cannot be argued against as a stipulative determination of rationality. It should not be thought, however, that the second ability, the ability to give one's reasons, allows an animal to perform some second process of reasoning or theorizing unavailable to animals without the ability to communicate. One can by fiat define a rich man as any man with £100,000 and a yacht, but the second condition, like Bennett's second condition, does not do the sort of work one expects in a useful definition. Communication adds considerably to the raw material of reasoning - the information available - not only in allowing the sharing of information among reasoners, but in allowing new types of information to contribute, but communication does not allow any new rational process to go on. And, as I shall show in §38, giving one's reasons is not necessarily reporting what reasoning one has actually done.

Ryle has objected to my use of "reasoning" by arguing that the familiar business that both merits the name reasoning and normally gets it is the business of deploying or expounding an already achieved and formulated argument or theory - pedagogically, forensically, or among colleagues. If this is the best use of "reasoning" then of course animals do not reason, and of course few people ever reason - nothing is more obvious. But then the ability to reason has little to do with the question of what is normally called <u>rational</u> behaviour or action. The activity of interest to me (and Bennett) can perhaps better be called "operating rationally with information". Reasoning (in Ryle's sense) then involves operating rationally with information, but is not itself operating rationally with information. My point then becomes that reasoning is something that only communicators can do (including phonographs?), and they do it well or ill, depending on their abilities as public speakers, but reasoning is not anything requiring special new rational mechanisms, processes, events, acts, intuitions, or whatever.

36. The Rôle of Awareness in Reasoning.

The error in speaking of having "rules or universal statements" is the assumption that consciousness is an arena in which reasoning (in my sense, not Ryle's) is performed. Behind most accounts of reasoning, I believe, there lurks an idea or intuition or vision of consciousness or awareness as an arena into which are led propositions, logical operators, and universal rules. The logical operators, like drill majors, direct the propositions into proper marching order, subsuming particulars and classifying concepts according to the behests of the universal rules, and then produce out of thin air a concluding proposition to bring up the rear. The audience is the internal eye, which introspects the parade and then reports to the world at large. Of course this vision is perfectly silly, and no one believes in it, but parts of it have a way of creeping into more sophisticated discussions of reasoning.

The vision lurks whenever one talks of thinking with words. Feigl says, "Our thinking is essentially mediated by symbols"[1] and Sellars says "Conceptual thinking is not by

1 Op. cit., p. 460.

accident that which is communicated to others."[1] Both of these remarks are harmless enough, if cautiously interpreted, but it is only a short step from them to: "When we think with words we often use them, like counters, to stand for things or ideas, in order to save ourselves the trouble of having to think every time about the things or ideas themselves."[2] Words, as words written or spoken or "entertained in awareness", are logically inert; they cannot react with one another, mate to produce conclusions, or subsume one another, any more than the picture of the Mona Lisa can be hungry or generous. I can say "I judge that ... ", but that is no more judging than saying "I walk" is walking. One must not be misled by such words as "deduce", "therefore", and "subsume" into assuming that reasoning is a manipulation of words that goes on <u>inside</u> (on the speaking side of) the awareness line.[3]

We are aware of the content of those signals that cross the awareness line. But when a signal has crossed the awareness line the only power it has is to trigger speech. It or its content or part of its content cannot combine logically with other signals to produce conclusion signals. The whole idea is misplaced. If a metaphor for awareness

---

1 <u>Science, Perception and Reality</u>, 1963, p. 17.
2 <u>Brain, Mind, Perception and Science</u>, 1951, p. 30.
3 Cf. Ryle, <u>The Concept of Mind</u>, Ch. IX.

is wanted, it can be likened to a broadcasting studio, not an arena. The action goes on outside the studio and the outcome is reported or broadcast; the parade itself does not occur in the studio. Ryle says, "A theorist cannot tell things, before he can tell them."[1] This means: one is aware only of the results of reasoning, not the process. The process is not the content of some signal, but the actual operations of neurones, and hence is not the sort of thing one could be aware of, barring self-surgery.

Consider the mathematician who does a problem in his head without even saying the steps to himself. Should we say he did the problem without <u>thinking</u>? He can tie his shoe without thinking, so why not solve a problem without thinking? Thinking can be analyzed to mean saying to one-self or having an awareness (thinking thoughts), or it can be analyzed to mean having neural activity going on. If the latter, then one does not tie one's shoe without think-ing. If the former, one could perfectly well solve a prob-lem without thinking. There is no middle ground. Sellars says at one point that to be able to think one must be able to criticize, but what does he mean by "think"?[2] Does he mean reasoning, or having an awareness? I suggest that he

---

1 *Ibid.*, p. 298.
2 *Op. cit.*, p. 6.

spiritual, or physical - it is. But he then proceeds to discuss "mental acts" or "subsumption" or "conceptualizing" in ways that are not neutral with respect to the various possibilities. The avowed <u>epoché</u> has not clearly succeeded, and the type of difficulties encountered in this supposedly neutral sphere can often be cleared up by returning, at least for a while, to the questions about just how these phenomena are substantiated.

To return to the mathematician, what additional aid would it be to him to say the steps of the problem to himself? Saying the steps must involve activity on the far side of the awareness line, since a signal is produced that crosses the awareness line. Saying the steps would have the effect of boosting signals with the same content as that which enters awareness into the operating areas of the brain <u>outside</u> the awareness line. It is the same function as is served by looking at each item on a bill separately when adding it up. For information to activate the "reasoning" areas of the brain, it must be facilitated. One cannot add up a bill that just passes before the eyes. If someone has written "6 + 4 = ?" on a blackboard, I cannot say "ten" unless I pay attention through facilitation to the information. Practice would seem to have the effect of sensitizing the reasoning areas so that weaker signals can trigger results, and also the effect of allowing more steps to be carried out with one set of facilitated signals.

When a person is stumped on a problem and keeps saying to himself over and over a set of "premises" with no result, he is boosting these signals into the brain (unless he is inadvertently "blocking" them through inhibition as soon as he has boosted them, which probably often happens) in the hope of getting a response or product. If nothing comes, he is either boosting premises from which nothing follows for him (since he has no organizations for coping with them), or he has bitten off more than he can chew - no matter how much he boosts the signals no results come because there are too many steps, too much transformation to be handled at one time. Often we boost some information into the brain and out comes a result that is farther along than we can explain. We recognize it as true (it "strikes" us that this is the solution we were searching for) but we cannot report the intermediate steps that must be there in a logical presentation of our deduction. Recognizing the result as true is simply getting the result, and does not involve being able to present in the approved step-by-step form "how we arrived at it".

In ordinary discourse we usually leave out of our presentations (our reasoning, in Ryle's sense) many of the requisite steps. When I say "It only costs a pound, so it can't be a real antique" there are many "logical steps" left out of my presentation. Information which I do not mention contributed to the result: information about the shrewdness of antique dealers, the law of supply and demand, the going rates for antiques. It is clear that having this information stored in my brain was required for me to reach my conclusion, and yet I did not consciously think about any of it. None of it came to awareness, but it was efficacious nevertheless. A person lacking this unmentioned, "unconsidered" information would not have been able to produce the reasoning (in Ryle's sense): "It only costs a pound, so it can't be a real antique".

It does not seem to follow from this that the logical steps we write down by convention when reporting reasoning parallel distinct and separate operations or events in the brain, but only that the information used in each step must have contributed to the organizations that produced the result - unless we have "jumped to conclusions". Dividing cerebral activity into separate events or steps must be fairly arbitrary; it is for this reason that Ryle's embarrassing question rings absurd: "How many cognitive acts did he perform before breakfast, and what did it feel like to do them?".[1] The modes of presentation established for logical writing or speaking are thus artificial, but they serve to enumerate the elements of information required for rigorous reasoning (although of course a person's own organizations may not be absolutely rigorous), and they illustrate the limits of what can be rigorously produced from any collection of information by building up the collection in steps. Thus, although such verbs as "deduce", "negate", "conclude" and such nouns as "premise" and "lemma" refer to no simple neural events or objects, there is such a thing as logical operation in the brain. The notion of logical steps, moreover, is not entirely artificial. When we break down a difficult problem by boosting parts of it, bit by bit, through the cerebral mill, we are in effect taking small logical steps.

37. Thinking.

There is one more way in which reasoning might be reserved for human beings. When Sellars argues that to be

1 Op. cit., p. 292. Geach, in Mental Acts, says, "The exercise of a given concept in an act of judgment is not in general a definite, uniform sort of mental act; it does not even make sense to ask just how many concepts are exercised in a given judgment." (p. 15.) This, I should have thought, amounts to an admission that the use of the terms "act" and "concept" here is poor. Geach nevertheless goes on to use the terms.

able to think, one must be able to criticize, this could well be interpreted to mean that the hallmark of reasoning is the ability of the reasoner to be critical of his own methods and results. Then one could say that even though animals' behaviour is controlled by mechanisms of a logical nature, animals cannot reason since they cannot be critical of their behaviour or the genesis of its controls.

This argument hinges on the interpretation of "criticize". If "criticize" means "criticize verbally" or "criticize consciously" (in the sense related to awareness), then animals cannot criticize since they have no verbal apparatus, and no awareness line. But if criticizing is taken as merely correcting what is wrong, guarding against errors, adapting to situations, then animals and many machines - even fairly unsophisticated machines - can be critical of their operations. Does the dog that relearns some behaviour to adapt to new surroundings criticize his old ways? If so, then he can reason. If not, then man is the only reasoner, but only in virtue of his being the sole possessor of language. The point I wish to stress again is that doing this or that consciously (with awareness) does not bring to the activity in question anything more than a relationship to utterance. Consciousness is not a semi-divine world of knowledge and understanding; it is merely a gateway to speech.

All "thinking" is not reasoning. Many of the neural organizations in the brain produce their results in ways that are not amenable to description in terms of logical operations or arguments. Racking one's brains for rhymes is thinking or pondering, but the results are not conclusions. Solving anagrams and crossword puzzles is not logical, or at most only partly logical. Playing chess, on the other hand, involves a good deal of reasoning, a high proportion of neural operation that produces information by logically combining other information.

A logical activity par excellence is theory-building or theorizing, and yet only a small part of the mental activity or thinking involved is reasoning. By "theorizing" I mean only the activity of producing theories, and not some special form of reasoning - different, say, from reasoning out a chess move or adding a sum or figuring out that the article is not an antique. Bennett, however, if I understand him rightly, holds that there is a special mental process, theorizing, which differs from the boosting (often by trial-and-error) of information into the cerebral mill.[1] To paraphrase Hume, I have never been able to catch myself at this activity, and cannot imagine what it is like. If Bennett can theorize he can do some-

1 See Rationality, p. 118.

thing I cannot.[1]

When one begins to build a theory, the thoughts that spring to mind are seldom conclusions. More often they are suppositions, possibilities, probabilities, tempting ideas, and high-sounding phrases. They are produced apparently by neural organizations of a much looser sort than those that produce conclusions. Some no doubt are simple association mechanisms, similar to those that help to produce the metaphors of poetry; some perhaps produce their results by always twisting the input according to some rather obsessional theme, some by casting the input into the fashionable modes of philosophical expression of the moment. It is the output of these organizations that is then channelled back into the rigorous organizations so that reasoning to conclusions occurs. It is hard to imagine how else theories could ever be built. Certainly no one has ever really sat down with a collection of axioms and proceeded to deduce truths from a dead start. Eventually one may find an elegant way of making one's point so that it rests on just a few simple truths, but putting to-

1 Animals, of course, cannot build theories. They cannot play the violin either - and why should they do either? Man cannot fell trees with his teeth, and why should he? It is a matter of the relevance of the activity to the being's existence.

gether theories or arguments almost never consists of putting together, right from the beginning, an elegant deduction.

38. Reasons and Causes.

Elizabeth Anscombe, in her excellent book, Intention, points out that the notion of giving one's reasons is central to the explanation of intentional actions. Bound up with this notion are the problems of "mental causes";[1] giving reasons and giving causes for one's behaviour are different things, and how they are different is of great importance. Since Anscombe's account of intention will figure heavily in Chapter 11, it is important that certain points she makes, and certain confusions in her account, be discussed at this point.

She begins her analysis of mental causes by trying to isolate the class of things that we "know without observation". (p. 13.) As examples she gives (1) the position of one's limbs, (2) that one has responded with a knee-jerk when the doctor has tapped one's knee, (3) the causes of certain movements, such as jumping when startled. She excludes knowledge of the location of pains, since

1 "Cause" throughout is meant in the sense exemplified by "Striking the match caused the explosion".

although one _can say_ where a pain is, one cannot be right or wrong about this, hence one should not say he _knows_ this. These examples do not serve to delimit a very useful class, and she does not give us much else to go on.

She is apparently not talking about the class of things we are aware of, the class of introspective certainties, the class of messages that cross the awareness line. The location of pains falls in this class. Barring verbal error, one cannot misreport what crosses the awareness line. Yet such a report is informative, so perhaps for this reason one should say one can be _right_ in one's reports. If it is decided that one cannot be right where only verbal error can occur, then it simply follows that one can be informative in making an utterance without being right or wrong.

Since Anscombe is not talking about the class of things known _without interpretation_, things known simply by their arrival at awareness, then she must be talking about some part of the complementary class of things known only through interpretation, only through taking the arriving signal as veridically indicating something. The two classes can be exemplified by a few pairs of things known: one knows without interpretation that one's foot is in pain, and one interprets the arriving signal to this effect as indicating veridically that one's foot is injured; one knows without interpretation that one seems to see a tree, and one interprets this signal as indicating veridically that there is a tree before one's eyes.

What then can Anscombe mean by "known without observation"? It is only with observation that I know there is a tree in front of me, and my knowledge of this does not differ in evidential status from my knowledge of the position of my limbs or that I have jerked my knee. In all these cases the arriving signals must be accepted or interpreted as veridical; in all these cases the contents of signals serve as fallible evidence for what is said to be known. Surely Anscombe does not mean to delimit the class of things known without visual observation, since then everything a blind person could know about the world would fall into her category. Presumably she is trying to get at the class of things we know through proprioceptive signals, the feedback signals from muscles and joints, and from motor cells themselves. But proprioceptive signals do not reveal the causes of movements, but only the occurrence of movements, and this only through interpretation, only if the signals are accepted as veridical. Knowing the position of my fingers does not differ in certainty or evidential basis from knowing that I am holding a cigarette in my fingers.

Anscombe saysthat the membership condition for the class of things known without observation is that nothing <u>shows</u> us that these things are so. But the only things that meet this condition she rules out, and none of her examples meet her condition, a hopeless muddle. She thinks that although one can be wrong about the position of one's limbs,

> nothing <u>shews</u> him the position of his limbs; it is not as if he were going by a tingle in his knee, which is the sign that it is bent or straight. Where we can speak of separately describable sensations, having which is in some sense our criterion for saying something, then we can speak of observing that thing; but that is not generally so when we know the position of our limbs. (p. 13.)

This is simply wrong. There is something of exactly the same functional sort as a tingle in the knee that tells us. There is no <u>tingle</u> any more than there is an itch or tickle on the retina when we see things, but there are signals that do show us the position of our limbs. The signals show us because they can be non-veridical and hence can fool us, in a way that pain-signals, taken as pain-signals and not as injury-signals, cannot. Anscombe claims: "You would know with your eyes shut that you had kicked when the doctor tapped your knee, but cannot identify a sensation by which you know it." (p. 15.) But you can; it is the sensation (in the ordinary sense) of kicking, which may or may not be caused by actually kicking. She claims such a sensation is not separable from the description of the action, whereas "the sensation of going down in a lift" is; i.e., one can have the latter when not going down in a lift. But this will not do. I can have the sensation of kicking when I am not kicking; it is just not as common an experience as having the sensation of going down in a lift when one is not going down in a lift.[1]

Proprioceptive signals tell us about situations in our limbs and other active parts; as such they can be wrong, if there is a fault in the circuits. The "safe" report is: "I _seem_ to have crooked my elbow." Pain signals are no different. We simply use them at their face value. We _could_ speak of them in the same way we speak of proprioceptive signals, as reporting injuries, and then the "safe" report would be "I _seem_ to have an injury in my foot" which

---

1 See also, _Intention_, pp. 49-50. "If a man says that his leg is bent when it is lying straight out, it would be incorrect to say that he had misjudged an inner kinaesthetic appearance as an appearance of his leg bent, when in fact what was appearing to him was his leg stretched out." Of course it would be incorrect to say that, but that is not what one would say if the man were mistaken. One would say he had judged an inner kinaesthetic appearance (a kinaesthetic signal crossing the awareness line) of his leg bent _as veridical_. If one is mistaken, it is the appearance or signal that is faulty, and accepting it as correct makes for the mistake; it does not make sense to speak of mistaking one appearance for another - and not just because it does not usually pay to speak theoretically of appearances.

is functionally or evidentially the same as "I have a pain in my foot". The ordinary pain-idiom has a built in "seem" operator. All this means is that we have chosen to tie the word "pain" to the incoming signal, and not to the injury the signal normally reports.

Anscombe is confused then about proprioception, but a more important muddle is her claim that one knows the causes of certain movements without observation. Now whatever she may mean by "without observation", it is perfectly clear that knowledge of causation is always inferential or through interpretation, and thus on an evidential par with the best examples of knowledge with observation.

Consider her example: "'Why did you jump back suddenly like that?' 'The leap and loud bark of that crocodile made me jump." (I am not saying I did not observe the crocodile barking; but I did not observe that making me jump.)" (p. 15.) What does she mean by saying she did not observe that making her jump? It is obvious that she did not see it making her jump the way one can see one billiard ball making another move. Presumably she means that nothing shows her that the crocodile's bark made her jump, and this suggests that she thinks her knowledge of causation here is non-inferential or immediately apprehended. But immediate apprehension of causation is an extravagant notion, and a most uncharacteristic metaphysical lapse on Anscombe's part. It is true that there is something I know about my jump that another person cannot know. I know that I saw (or seemed to see) the crocodile leap and bark, and I know by proprioception that I jumped (or seemed to jump), and so far as I know nothing else entered into the situation. I know I did not think to myself, just before the crocodile barked, "I think I'll just jump back for the fun of it", and I know I am not afflicted with some malady that makes me jump every now and then. So I assume that it was the sight of the barking, leaping crocodile that made me jump, and this assumption is fairly safe. But I do not have non-inferential or immediate knowledge of the cause of my jump. And of course it is only contingent that another person cannot know what I do about my jump. Neurologists might some day know just as well - in fact better - what caused my jump. I have no access, private or otherwise, to my neural activity, but only to my awareness and the succession of messages arriving there; having no other explanations of the jump, and having seen others jump when presented with sudden, strange sights, I infer that the startling sight caused the jump.

Suppose I am crying, and someone asks why. I say "because Smith just died". I am assuming, again, that there is a causal relation between learning the sad news and crying, since the occasion is similar to other occasions on which people have cried. The regularity with which the receipt of sad news is followed by crying suggests that there is a causal relation between the two, and neurologists may soon help to confirm this hypothesis. But all that I, the crier, may know that another person may not, is that in <u>this</u> case nothing else of conceivable relevance, such as an onion or directions in the script: "cry here", has entered into the case.

As Anscombe points out, her view that there is knowledge-without-observation of certain causes cannot be accommodated to Hume's explanation of causality. (p. 16.) Indeed it cannot, and it is wrong. Reporting the "mental" causes of certain actions, like reporting their very occurrence and like reporting the position of the limbs, is fallible and inferential - unlike reporting the existence and location of pains, or the occurrence of other "events in awareness".

Reporting reasons falls into the same category as reporting causes; it is not foolproof like reporting pains or experiences. When one is asked to give one's reasons for an act, one is asked to give the reasons that actually worked, that led to, or determined the act, and not just any plausible reasons that come to mind. There is a tradition in philosophy calling for the strict separation of reasons and causes, and Anscombe is in the tradition. She wants to say that when there is a reason for an action there is not a cause, and _vice versa_. Yet it is quite clear that, when one is asked to give one's reasons for an action, what is wanted is an account of the reasoning that made up part of the causal chain producing the action. We do not have direct access to our reasoning operations, which are bits of neural activity, but only to the train of thoughts or conclusions that may cross the awareness line. And even our awareness of such a train of thought does not ensure that the reasoning involved has played or is playing a part in the direction of our actions.

I may know that I have thought "I am very upset, and a drink would calm my nerves, so I'll have a drink", and I may then have a drink. My belief may be that the reasoning _behind_ this thought (for thinking the thought is not itself reasoning) caused or determined my behaviour. I may believe, in other words, that I would not have had a drink if I had not reasoned out the beneficial effect of a drink on my nerves. But I cannot know this certainly. If I am not a compulsive drinker always looking for a good excuse for a medicinal nip, my belief is probably well-grounded, but not certain. If someone wants to know _my_ reason for taking the drink, it is not enough, finally, for me to give him the reasoning that "occurred to me", for this may not have had anything to do with determining my behaviour.

If someone asks me for the sum of 19 and 32, and I reply after a moment "51", and am asked "Why did you say '51' and not some other number?", I will answer that I wanted to give the right answer, and then describe the reasoning (the adding up) that led me to the right answer. Here my account of my reasons is surer, simply because it is hard to imagine what else could have prompted me to say "51" except this. One does not go around announcing two-digit numbers at random. It would be very strange if my saying "51" at that moment were not in fact a reasoned response to the question. For that reason the question "Why?" here is silly, and would hardly ever be asked, since the answer is so obvious. But still, I have an evidential edge on my interlocutor; I know what thoughts ("two and nine are eleven, carry the one ... ") were thrown up by the reasoning process before I issued my answer. But it should be noticed that if I am a good mathematician, I may produce the answer without any such accompaniment in awareness.

If I, without stopping to think, pull a child away from a fire and am asked to give my reasons, I may say with a high degree of certainty, "because I saw he would soon be burned", and in this case I am not relying on any remembered thoughts that may have run through my head - for none did - and my knowledge that this is the correct reason is based on my knowledge that I did not think "I'm going to kidnap that child" or "let's put baby in his crib for a while". Lacking any evidence for exotic explanations, I infer that I recognized the danger and then acted on this recognition - which is the obvious explanation. But unlike the other cases, I have no positive evidence for this account.

In some cases, I may say that I do not know for what reasons I did something, but this does not always mean that I did not act for some reason. Sometimes, the question "Why?" has no application in terms of reasons, but just of causes: "Why are you blinking your eye?".[1] It should not be supposed that just because some actions have causes but not reasons, others have reasons but not causes. Reasoning, when there is reasoning, is part of the causal chain that produces the action.

But Anscombe, if I understand her, does not wish to speak of there being causes when there are reasons. She says:

---

1 Cf. Ibid., pp. 25-7.

> ... how would one distinguish between cause and reason in such a case as having hung one's hat on a peg because one's host said "Hang up your hat on that peg"? ... Roughly speaking - if one were forced to go on with the distinction - the more the action is described as a mere response, the more inclined one would be to the word "cause"; while the more it is described as a response to something as having a significance that is dwelt on by the agent in his account, or as a response surrounded with thoughts and questions, the more inclined one would be to use the word "reason". But in very many cases the distinction would have no point. (pp. 23-4.)[1]

There is, of course, causation in both cases. Where one ponders, the path from heard request to behaviour is simply less direct and less swift. There is reasoning going on even when one "unthinkingly" hangs one's hat on the peg; the behaviour is appropriate to the stimulation because it is mediated by organizations established by stored information - about manners, pegs, hats, and so forth. The "unthinking" response is leagues beyond the Pavlovian conditioned response (people are not trained to hang their hats up on the hearing of verbal cues), and leagues more beyond the knee-jerk, which is hereditary. How very strange it would be that a person should hang up his hat in response to a verbal cue, unless it were a swiftly reasoned response.

---

1 See also pp. 10-11 and 34, and Bennett, op. cit., p. 44.

39. <u>Two Misgivings Considered</u>.

I have said that the brain performs operations when one is said to be reasoning and that we are not aware of these operations, but only of the results. I have also argued that although we can be aware of the results of our reasoning, we cannot know with certainty that these seeming results are caused by actually occurring reasoning, or by any process that is having an effect on the direction of our behaviour. There is no sure way, in other words, to distinguish rationalizing from efficacious reasoning.

It may be objected that since I can never know certainly that my reasoning, or the results of my reasoning, direct my behaviour, then I do not know if my reasoning or its results <u>ever</u> direct my behaviour. If so, then why do I continue to reason if I do not know that it has any effect?[1] But I do know that my reasoning has effect, or rather, my brain "knows" this, since the operations of the brain depend for their existence on their proven utility. Although there is certainly good reason to believe the psychoanalysts when they claim that certain long-range goals, certain higher types of behaviour, are caused by irrational mechanisms, all human behaviour cannot be irrational. We often act in

---

1 Cf. Anscombe, <u>op. cit</u>. p. 52, for a refutation of similar arguments.

ways that we fully recognize as irrational or non-rational,[1] even when we ought to be rational, but these occasions are rare. To imagine a person whose behaviour is completely irrational is to imagine an unseeing, inactive slug, or a flailing monstrosity. We can imagine a demented poet who tries to determine his behaviour not by reason but by rhyme - by the operation of his mechanisms of poetic association: "Why are you baking a cake?" "Because the moon is on the lake"; "Why are you chewing the spoon?" "Because I hear a mournful tune". But if he succeeds in baking a cake, or even in giving semblances of answers to our questions, he must be doing a fair amount of reasoning.

Finally, there may be some who will cling to the notion that physicalism is an out and out absurdity, and they will put their "embarrassing question": "But how could a brain, a hunk of mere organic stuff, a collection of pulsating cells, actually reason?" One is meant at this point to conjure up images of the noble Descartes, the sublime Einstein, and others, and then laugh the notion of the reasoning brain out of court. Several fires produce this smokescreen. First, one need not say, if it is found to be unsettling, that the brain reasons. One can say that people reason -

---

1 By "irrational" I mean against reason, mad; by "non-rational" I mean having nothing to do with reasoning.

with their brains; the brain can no more reason than the feet can flee or the hand can paint a masterpiece. Second, reasoning means many things, depending on the context. The embarrassing question is somewhat like the question: "How can two small cubes of ivory, with different numbers on each face, bring ruin and tragedy to a man's life?" A world of behaviour, institution, and description has accrued over the years to the action of dice, and similarly, centuries of discourse, ordinary and theoretical, about sublime reason and the mental life have produced a rich mystique about reasoning, but the nugget of physical process behind all the talk lies in the brain, just as the physical difference that triggers the gambler's sad fate lies in the position of two dice on the table.

CHAPTER 11.

Intention.

40. Awareness and Action.

The concept of intention is not part of the scientific panoply; very little has been said on the subject by psychologists, and next to nothing by neurologists. In fact, neurologists tend to sidestep the issue by prefacing their remarks on the activity of the motor nerves with the standard expression: "And although we can say nothing about how motor activity is initiated ... ". This certainly suggests that at the upstream end of the neural river there will be a small sign: "This far, no farther", and the mystery of "initiation" will remain forever unsolved. The mystery of intention is something of a last bulwark for those who wish to maintain the notion of a vital spark of non-physical something at the control console of the body.

Cybernetics has an answer to the mystery. It simply denies that there is a fundamental distinction to be made between "unintentional" actions (such as knee-jerks) and "intentional" actions. All human behaviour, so goes the theory, is merely responses to stimuli, and "acts" differ only in the degree to which the response is mediated by neural organizations established by the influence of past experience.

But this does not and should not satisfy. Rightly or wrongly we place great importance on whether an act is intentional or unintentional - in law, for instance - and whatever criteria we may have for distinguishing between the two should be uncovered. If it turns out that the difference is purely a matter of linguistic institution or a vestige of a creed outworn, with no actual functional difference behind the application of the terms, then our placing so much importance on the distinction is surely wrong - both unscientific and morally wrong. But it would be very strange if this turned out to be the case. It is hard to imagine how a baseless or mistaken "language game" of such life-and-death importance could have arisen and survived.

The problem of determining these criteria has been greatly complicated by the infiltration of moral and quasi-religious questions about freedom of the will, responsibility, and moral laws, and by the numerous false starts that have attempted to explain the act of willing, the nature of "initiation", and the operation of the mind-body causal link. Miss Anscombe, in Intention, isolates the problem of determining the criteria for the application of "intentional" by examining the operation of the question "Why?" asked about certain acts.

It is not bodily motions, she points out, but motions <u>under particular descriptions</u> that are intentional or unintentional. I may be sawing a plank, and it may be one of Smith's oak planks, so that I am sawing a plank, sawing one of Smith's planks, and sawing an oak plank. But these are not different motions; in doing all three at once I am not performing three separate feats of motion. The action of sawing the plank (the motions considered under that description) can be intentional, while the action of sawing one of Smith's planks (the same motions under a different description) is not.

In general, Miss Anscombe shows, for an action (by which is always meant a motion under a particular description) to be intentional, it is necessary that the actor be <u>aware</u> of the action under that description. If I am typing, and someone asks me "Why are you tapping out the rhythm of 'Rule Britannia'?", and I reply, "Oh, am I doing that? I was not aware of it", I show (if I am truthful) that the action of tapping out the rhythm was unintentional. Not being aware of tapping out the rhythm, I cannot be doing it intentionally. Awareness, however, is not a sufficient condition. I may be aware that I am doing one thing in the course of doing something else, and yet not be doing the former intentionally. I may notice that I happen to be neatly not stepping on the cracks of the pavement, and yet not be intentionally avoiding them; they just happen to match my normal stride. I may, in fact, pay particular attention to this phenomenon to see how long I can keep to my normal stride before my foot lands on a crack, and in such a case I am still not avoiding the cracks intentionally, but only keeping to my normal stride intentionally - however much I <u>want</u> to keep on avoiding the cracks.

Anscombe takes account of this by distinguishing intentional actions as members of a subclass of the class of actions of which one is aware: the class of one's actions of which one is aware <u>without observation</u>. That is, one denies that an action is intentional if one says "I only <u>observed</u> that I was doing that". (pp. 24-5.) So supposedly we have some sort of direct awareness or knowledge without observation of our intentional actions. But as I showed in the last chapter, Anscombe's concept of knowledge without observation is confused. She tries to establish a class of things known both without anything <u>showing</u> one that these things are so, and also without our knowledge or awareness of these things being incorrigible. In this she has tried to have her cake and eat it too. The only time knowing (or reporting, which comes to the same thing) is incorrigible, save verbal errors, is when what counts as the thing known is just the content of a signal arriving at the awareness line, and nothing shows us that we are aware of what we are aware of. (See Chapters 7, 8 and 10.) The way out of this dilemma is to drop one of Anscombe's two requirements, and the one to drop, quite clearly, is the requirement that nothing must show us what we know. This can be replaced by the requirement that what shows us is not a signal from any of our sense organs, including stretch detectors in muscles and other kinaesthetic organs.

What emerges from this emendation of her account is a perfectly straightforward description, in the terms established in previous chapters, of the class of actions "known without observation". An action falls into this class if a signal with a content in some way descriptive of the action crosses the awareness line, and this signal is not a proprioceptive signal from muscles or joints or a signal from the sense organs, or produced by either of these. Now could any signals fit this description?

In the chapter on content it was pointed out that high level motor signals, the signals that trigger and control the appropriate concatenations of motions, or actions, in other words, can be seen as both like orders and like pushing buttons. The sophisticated function of these signals allows them to be considered as orders having a content descriptive of the action being directed. Now if a general issues an order for his army to retreat, and this order is successfully implemented, the general's order amounts to a description of what happens. Similarly, if the high level motor signal that triggers and directs an action were to send its signal also across the awareness line, the utterance of its content would be a description of the action that takes place, barring failure of the subsidiary nervous activity, the muscles themselves, or prevention by the environment. And whether or not one utters this description, if the message were to cross the awareness line one would be aware of its content, which would be a particular description internally related - in that it is the content of the triggering "order" - to the motions caused. One's awareness of this description, though not of the actual occurrence of the action, would be direct, and utterance of it would be subject to only verbal error. One's awareness of this message would <u>show</u> one that he had "ordered" the action described.

Does this account upset the notion of signal content established earlier? Content, it will be remembered, was to be assigned on the basis of both stimulus conditions and continuation of function. The branch signal that would cross the awareness line in this case would not have the function of controlling the action described so it should perhaps be called a different signal with a different content. This is, however, no problem. If the stimulus conditions for the branch signal are the same, i.e. it fires only when the signal fires that triggers and directs the action, then it only differs in having the function of making a certain utterance possible. In that case, its content would be more "declarative" and less "imperative", more to the effect of "left hand now ordered to grasp the lever" than "left hand: grasp the lever!"

I am positing the occurrence of a certain type of neural activity: the branching of high level motor signals across what I have called the awareness line, and I must admit that I have no physiological evidence for the existence of such activity. But there is also no physiological evidence against such activity, and there is an abundance of anatomical evidence to show that the relevant parts of the brain are interconnected richly enough for such activity to occur. The activity is posited on the evidence, so well adduced by Anscombe, that we sometimes can report what we are doing without relying on our sense organs, or on "observation". Also, there is no denying that the posited activity fits nicely into the account of neural activity already proposed to deal with quite different matters.

An example is in order to bear out the coherence of this account. Suppose a man with a hammer decides to finish driving a nail into a door. Suppose that, for example, he happens to see the nail sticking out, and this visual input, after running through various cerebral mills, produces a signal that triggers the motor activity of driving in the nail. This description eliminates the word "decide" which could be interpreted as begging the question of whether the action was intentional. The high level motor signal sends its message across the awareness line, and the man thinks (says to himself) something like "about time I finished driving in that nail". It is not necessary, of course, that the inhibition that prevents him from actually uttering the message occur just prior to motion of the lips, lungs, and so forth. He need not go through the temporal process of saying the words to himself. (See Chapter 7.)

So his thought has the content,[1] approximately, "I'll drive in that nail", and on this account, the high level

1 More rigorously: he is able to utter a message having the content "p". This ability is to be construed thus: he will utter words to the effect of p unless there is inhibition of the implementary activity between awareness line and vocal muscles.

motor signal that triggers the nervous activity that directs the action of driving in the nail also produces the signal that crosses the awareness line.

Suppose instead that the man is prompted to make a dent in the wood next to the nail - for some arcane reason. A different high level signal produces this action, and the neural activity for this action differs from the neural activity that would occur if he accidentally missed the nail and made the dent. Missing the nail accidentally would be the result of a malfunction of the implementing signals or of weakened muscles, or it could be the result of poor programming of motor neurones in the first place, or as we usually say, the result of poor coordination.

Anscombe gives an example where the utterance of the description and the action do not match:

> ... I say to myself 'Now I press Button A' - pressing Button B - a thing which can certainly happen... And here, to use Theophrastus' expression again, the mistake is not one of judgment but of performance. That is, we do not say: What you said was a mistake, because it was supposed to describe what you did and did not describe it, but: What you did was a mistake, because it was not in accordance with what you said. (Op. cit., p. 57.)

As she points out, this is just like obeying an order wrong, which is not a case of disobedience, but of malfunction. Of course, the malfunction can occur in the implementing of either the action or the verbal utterance. I may say

"Now I press Button A" and mean to have said "Now I press Button B". It is even possible for malfunction to occur between the firing of the high level motor signals and the awareness line, so that I am <u>aware</u> of "ordering" "Press Button A", while the brain has actually ordered "Press Button B". In this case, I am not mistaken in what I am aware of (I cannot be), and I have not made a mistake in carrying out the neural order. I am not mistaken about anything unless I assume - which I am likely to do - that the signal to the effect that I am ordering "Press Button A" is veridical.

I should like to call this knowledge <u>via</u> motor signal "motor knowledge";[1] it is certainly what Anscombe was after with her term "knowledge without observation" (p. 53 and especially p. 82). When I perform an intentional action I have motor knowledge of what I am doing. I am able to utter descriptions of my actions unmediated by observation with sense organs. But I do not know certainly that the signal that produces this description in awareness <u>causes</u> the action. It usually does, but learning this is making an empirical discovery, learning that there is a regular

---

1 I am accepting that one can <u>know</u> what one is aware of. There are certain qualifications and elaborations to this, which will be dealt with in the next chapter.

correlation between these things I can say and the actions I observe myself performing. I have certain, unmediated awareness of a description of an action, but not of the action itself, or of any causal bond between description and action. I also have knowledge of my action through proprioception and perception, but this is of course mediated, observational, uncertain knowledge.

The fact that our knowledge of what we are doing intentionally is twofold in this way is somewhat hidden from the introspector by the functional interdependence of these two "modes" of knowing. Neurologists have determined that without proprioceptive signals all but the most simple motions are impossible to perform. The organizations of lower level motor neurones that translate high level signals into complex motions depend on a steady flow of feedback. Without it the motor complex stalls. For example, when one awakes to find one's arm "asleep", one is unable to do more than flop it around. This is apparently not so much because the motor fibres are blocked, but because the proprioceptive fibres are blocked, and the motor organizations depend on the proprioception for coordination.

When, to use one of Anscombe's examples (p. 53), I write words on the blackboard with my eyes closed, my knowledge of what I am doing is twofold: I have motor knowledge of the high level "order", and I know by proprioception that the action is occurring properly. Without the proprioception the act would be impossible. Opening by eyes and looking adds to the feedback. This additional feedback is virtually essential for some actions, such as drawing a cow, but not essential for many others, such as signing one's name and playing the piano, even though it is initially invaluable in coordinating these actions. So it is impossible to isolate an interesting action, as Anscombe tries, where one's only knowledge is motor knowledge. Her analogy of a man directing the construction of a house by remote control with no feedback thus has no direct counterpart in human actions, but this is, of course, only contingently so.

So Anscombe rightly sees that there are two types of knowing, and describes how they are relevant to the problem of intentional action, but her account is marred by several factual errors. Motor knowledge <u>is</u> knowledge where something (the signal whose message I am aware of) <u>shows</u> me something (that I am ordering a certain action), and this is so in that I can be mistaken, if the sign by which I know is not veridical. In this respect motor knowledge does not differ from proprioceptive knowledge or visual or auditory knowledge. All are mediated by signals, which, although they cannot be misidentified, can be faulty in the first place, the product of a malfunction between sense organ and awareness line. These cases all differ from the case of awareness of the location of a pain, where the arrival of the pain signal at the awareness line, whether veridical or not, simply constitutes having a pain.

With these minor points cleared up, Anscombe's ideas on action and awareness mesh nicely with the theory so far developed, and yet Anscombe explicitly rejects physicalistic explanations of intention. In fact, she offers a proof "that an action is not called 'intentional' in virtue of any feature which exists when it is performed." This is surely a strange notion to set out to prove. It seems quite obvious to me that, for example, there must be differences in the neural activity of the man who slips on a banana peel accidentally and the clown who slips _intentionally_.[1] And similarly for the man who intentionally makes a dent in the wall instead of simply missing the nail. One wonders if Anscombe wants to prove this, or is simply driven to it because she cannot see what feature it could possibly be.

> Let us suppose that there is such a feature, and let us call it '_I_'. Now the intentional character of the action cannot be asserted without

1 Cf. Ryle, _op. cit_., p. 33.

> giving the description under which it is intentional, since the same action can be intentional under one description and unintentional under another. It is however something actually done that is intentional, if there is an intentional action at all. A man no doubt contracts certain muscles in picking up a hammer; but it would generally be false to call his contraction of muscles the intentional act that he performed. This does not mean that his contraction of muscles was unintentional. Let us call it 'preintentional'. Are we to say that I, which is supposed to be the feature in virtue of which what he does is an intentional action, is something which accompanies a preintentional action, or movement of his body? If so, then the preintentional movement + I guarantees that an intentional action is performed: but which one? Clearly our symbol 'I' must be interpreted as a description, or as having an internal relation to a description, of an action. But nothing about the man considered by himself in the moment of contracting his muscles, and nothing in the contraction of the muscles, can possibly determine the content of that description; which therefore may be any one, if we are merely considering what can be determined about the man by himself in the moment. (pp. 28-9.)

The last sentence is, I hold, simply wrong. There is something about the man at the moment he contracts his muscles that determines the content of the description: the triggering high level signal. The signals that trigger action are double-edged in a way Anscombe assumes nothing can be: they both cause the action and describe it; as quasi-orders, they have descriptive content. This account holds just as well for the radio signals that send "orders" to satellites, as it does for my proposed neural signals. Anscombe has done a curious thing. She has considered acting man as a proper subject of physiology, for she has discussed the contraction of his muscles and concluded that nothing in the physiology of his muscle movement could determine the content of the description, and yet she has completely ignored the one field of physiology, neurophysiology, that obviously could have relevance to intentional action. She does not say: nothing in the <u>nervous activity</u> that <u>directs</u> muscular motion could determine the content of the description of the action. Has she simply overlooked the possibility?

Philosophers seldom commit themselves to views that are directly confirmable or disconfirmable by scientific evidence, in this way protecting themselves from the onward rush of knowledge, but here is one instance where scientific evidence <u>could</u> disconfirm a philosophical hypothesis. I cannot say, of course, that there is at the moment any strong disconfirmatory evidence, for my description of neural activity is only hypothesis, far ahead of any substantiating evidence. Before any evidence could disconfirm Anscombe's view, my mode of description in terms of content and signal function would have to be borne out by a wealth of data not now available, and more important, by its proving to be a useful and unmisleading way of describing nervous activity. All that can be said at the moment, then, is that Anscombe offers no argument; the last sentence is far from being obviously true; and I offer a contradictory theory with some degree of evidential backing and internal coherence.

41. Reasons and Intentional Action.

The objection can be raised to the account given in the previous section that although a high level motor signal might have the content: "Saw this plank", it would be foolish to suggest that a different motor signal would be needed to have the content: "Saw Smith's plank" or "Saw the oak plank". That is, although the idea of description by motor signal works well enough for pushing buttons and driving nails, one cannot expect a motor signal to have the content "Cook Smith's goose" when directing the action, say, of signing a formal complaint about Smith. One may intentionally be cooking Smith's goose in signing the complaint; if asked "What are you doing?" one may well and truly reply "I'm getting Smith good and proper", and that is then the description of the intentional action, but surely no motor signal will have that for its content. This is where reasoning comes in.

The conditions for intentional actions outlined in the previous section are necessary, but not sufficient. One may have motor knowledge of an action and still not be doing it intentionally if one's answer to the question "Why are you x-ing?" is "no particular reason" or "I don't know; I was just doodling". As Anscombe points out, such an answer is not a rejection of the question, as "I was not aware I was doing that" is. "The question is not refused application because the answer to it says that there is no reason, any more than the question how much money I have in my pocket is refused application by the answer 'None'." (p. 25.) The final requirement for an action to be intentional is that there must be a reason that can be given by the actor for the action. Where there is no reason in the offing, as in doodling, Anscombe would call the action voluntary, but not intentional, and I see no reason not to follow her usage. Other actions may qualify as voluntary, but at least all actions of which one has motor knowledge but can offer no reasons for doing are voluntary.

Reason giving, however, is not a foolproof activity. (See Chapter 10.) Even when pondering occurs, so that we are aware of the input and output of reasoning - but not of the actual reasoning process - we cannot know with certainty that what we offer is reasoning and not just rationalization.

And when "conscious reasoning" has not occurred, our reason giving is simply conjecture, although often highly reliable conjecture. It should be noted too that although one can utter what one is aware of infallibly - save verbal slips - this infallibility disappears when one is reporting what one was aware of at some time in the past. In such cases one utters infallibly the memory one is aware of, but the memory itself may not be reliable.

Even when one has pondered, one may only have pondered the last few steps in the long process of turning input rationally into behavioural output. The "Why?" routine brings this out: "Why are you sawing the plank?", "I'm making a table", "Why are you making a table?", "Because we need one", "Why do you need one?", "To put our food on", "Why put your food on a table?", "Just because, that's why". After the initial answer or two, what follows is largely conjecture or fabrication, not based on any pondering that one has ever done. It is true or false, since either the information cited has contributed indirectly to one's behaviour or not, but it often will not be recognized as true or false, but just as likely, by the responder. The repetition of "Why?" is supposed to have the effect of probing deeper and deeper into stored information and rational methods of the actor, but once the responder has reported, fallibly, what thoughts he was initially aware of that might have contributed to his direction of behaviour, the subsequent answers are merely parts of his own personal theory of motivation. If he happens to be right then there are in fact efficacious neural organizations in his brain that contribute to the direction of his behaviour along the lines indicated by his theory, but he has no foolproof way, and even no very reliable way, of checking his theory. But it still makes sense to ask the "Why?" question. He has his memories of past behaviour and the thinking that accompanied it, and this is certainly relevant information not held by the questioner. The actor is simply empirically better acquainted with his own style of behaviour than anyone else is, although he may not be particularly perceptive or critical about his own behaviour.

To return to the objection raised at the beginning of this section, it becomes clear that the action of signing the complaint is intentional in virtue of the reason for which it is done: to cook Smith's goose. The description of the action as cooking Smith's goose or getting Smith good and proper is a description that the actor can be aware of, if not as the content of a motor signal, then as the content of a reasoning signal. That is, the motor signal that triggers and directs the activity of signing the complaint can be seen as part of the implementation of the signal with the content: "fix Smith good and proper", just as the low level motor neurone firings that stimulate the individual muscle motions are part of the implementation of the signal with the content: "sign the complaint".

The distinction between motor signals and reasoning signals is not a clear cut one. No end is served by trying to pin down a particular point where the brain stops processing data or deliberating on it and starts issuing motor commands. The description that correctly fits the motion _qua_ intentional action depends on which signals at which level cross the awareness line. Closing the hand around the pen and proceeding to swing the pen across the paper in a certain complicated way is almost never the intentional action performed. Signing a document almost always is. If when one signs a document one is also aware of certain _reasons_ for signing the document, e.g., in order to fix Smith, or in order to rid the neighbourhood of a nuisance; or if one is also aware of certain wider _descriptions_ of what one is doing, e.g., fixing Smith, or ridding the neighbourhood of a nuisance, then one is also intentionally doing these in signing the document. Does one move the pen _in order to_ sign the document or _in signing_ the document; should one say that he signed the document in order to get

Smith, or should one say that in signing the document he was getting Smith? Nothing seems to hinge on the alternative between reason and action here. One starts a car by turning on the ignition, or one turns on the ignition in order to start the car. One pushes the pen in order to sign the document, or by pushing the pen one signs the document.

Thus the demand for reasons for intentional actions is not a demand with fixed requirements, since there is no fixed length for the reasoning one must give, and no fixed description from which to start. Sawing a plank is intentional if the question "Why are you sawing the plank?" is answered "I'm making a table". But if someone a bit more observant asks, "Why are you making a table?", this requires something in the way of further reasons. But then does the first question also require these further reasons? One is usually satisfied with the short answer, "I'm making a table"; it all depends on the interrogator's interests. It is not as if for an action to be intentional, the actor must be able to fill in satisfactorily a standard questionnaire requiring _n_ nested reasons for the action. Furthermore, as already pointed out, the actor's answers to the "Why?" questions soon become little more than conjecture.

42. <u>The Language of Intentions</u>.

I have said that the "Why?" routine is supposed to have the effect of probing into the neural organizations in the brain, and yet obviously the ordinary man has no such plan in mind when he asks "Why?". How could it be that giving reasons is making reports and conjectures about neural signals and organizations, when most people have no idea what is going on in the brain? This question is akin to the old standby: "How can a thought be a brain process? I know I have my thoughts, but I do not know I have brain processes." (See Chapter 2.)

The question here has no force, for two reasons. First, to adapt Smart's argument, I am not proposing that one *means* by "reason" or "intention" a particular neural state of affairs or event, any more than one would say the ancients *meant* electrical discharge by "lightning". The relationship proposed is a contingent, discovered relationship between neural activity and the application of the terms "reason" and "intention". And second, I am not proposing any identity of intentions or reasons with brain processes. I am not saying that the statement "My intention is X" can be altered to "My neural-organization-of-type-I is X", or that the question "What are your intentions?" can be altered to "What are your I-type neural organizations?" Again, it is the whole question, and the whole report, that has significance, as part of an area of verbal behaviour devolving on certain physical states of affairs. Behind the talk, I am saying, making the talk true or false, are "brain processes" and neural states, and nothing of a non-physical nature. That the ordinary word "intention" stands only as a counter or index in longer expressions, and does not refer at all to any of the events or states discussed is made abundantly clear in the attempted substitution: "Are your intentions honourable?". "Are your neural signals honourable?". The high level motor order signals come close to what one intuitively supposes an <u>intention</u> would be, but to identify the two, and then be bound to reply to such criticisms as "How can a neural signal be honourable?" is to commit a gratuitous solecism.

If ever the term "language game" is applicable to linguistic situations, it is applicable to the routine described by Anscombe of asking about reasons and intentions. When one is asked "Why are you doing X?" there are standard and accepted ways of responding, and if one does not abide by these ways of responding, it is very much like <u>not playing the game</u>. If someone asks me why I am typing and I reply "The horse is a sympathetic animal" I am simply not playing the game; if on the other hand I <u>lie</u> to him, there is - there must be - a physical feature in virtue of which I am lying, in virtue of which I am breaking the rules. Similarly, if it is my turn to bid in bridge and I say "straight flush, queen high" I am simply not playing the game; if I revoke on a trick, on the other hand, there is a physical feature in virtue of which I have broken the rules: the physical presence of a certain card in my hand.

There is no particular mystery in how we learn to play this language game without knowing the physical features behind the rules. Once one is programmed for language what one can say will fall in automatically with what others can say. Learning to speak is learning to produce utterances of whatever crosses the awareness line; the shape and form of the utterances is determined by what language one is learning to speak. And the genesis of the language game in the first place is similarly unmysterious. The utility of the particular verbal behaviour more or less ensures that it will develop, in one way or another. The internal structure of the utterances is of little importance; Italian has one word, "<u>perchè</u>" for both "why" and "because", but this does not show that Italians are somehow missing a distinction when they play the reason-giving language game.

43. Willing.

The account of intention that has been given includes no talk about volitions or willing. That is because, as Anscombe argues, the verb "to will" is a hoax. There are no such things as acts of will or volitions.

> People sometimes say that one can get one's arm to move by an act of will but not a matchbox; but if they mean 'Will a matchbox to move and it won't', the answer is 'If I will my arm to move in that way, it won't', and if they mean 'I can move my arm but not the matchbox' the answer is that I can move the matchbox - nothing easier. (p. 52.)

The idea that willing is some sort of radiation generated by gritting the teeth and saying, over and over again, "move, move, move" is hopeless. It arises, no doubt, from such experiences as lying in bed and saying to oneself "I must get up, I must get up; it's late. On the count of three: one, two three ... " until finally one gets up. The causal link in these cases has been debated at great length, for on the one hand, thinking these thoughts often seems to help or even cause the action, and yet on the other hand very often thinking the thoughts has no effect at all.

It is supposed, perhaps, that when thinking these thoughts does not work, one is just not thinking hard enough or with enough conviction, but these explanations are obvious dead ends. The "tone of voice" with which one says these things to oneself clearly does not make any difference, and what else can one do to simulate or bring about conviction? The facts of the matter - that sometimes the thoughts seem to help and sometimes not - suggest that thinking to oneself is merely an accompaniment or by-product of the actual business of determining action. It seems most likely, that is, that if I have the conviction that I must get up, I can say the words and will thereupon arise from bed - but I will arise equally well if I do not say the words to myself. And if I do not have the conviction that I must get up, I will not get up, whether or not I mutter exhortations to myself with great vehemence.

The theory of neural activity so far developed provides a plausible explanation of this phenomenon. Roughly, in order for the brain to initiate the activity of getting up, its input must be such that it outweighs, say, the pleasure of just lying in bed, the influence of stored information to the effect that getting out of bed is unpleasant, the input to the effect that the body is still tired, and so forth. As soon as the balance is tipped, the brain initiates the activity and one gets up. There is no need to suppose that once the balance is tipped something must recognize that the balance is tipped and then proceed to <u>will</u>

the act of getting up, any more than a spinning top must recognize that its gyrostatic force is no longer strong enough to balance it in order to "decide" to fall over.

When the balance has been tipped or is being tipped in the brain, messages that cross the awareness line to the effect that it is time to get up, there is much to be done, and so forth, are merely accompaniments to the "decision-making" of the brain. But when the balance has not been tipped, no amount of repetition of these messages, if in themselves they are not enough to tip the balance, will bring about the action.

The way, presumably, to tip the balance is to increase and facilitate the sort of input that would outweigh one's inertia. The input that would accomplish this would depend on the dominant organizations at the time. The information: "the British are coming!" would seldom serve to tip the balance. The information "it's time to get up" can tip the balance if the person has some reason to get up on time or some natural tendency to regularity and punctuality, which amounts to a "weakness" for this sort of input. But if the information "it's time to get up" is not sufficient to outweigh the inertia, no amount of feeding this information repeatedly into the relevant parts of the brain will tip the balance. It is not awareness or consciousness, however, that is producing this information, but the rest of the brain; awareness is simply the verbal outlet for the information.

Just as one person can devise input information to stimulate another person - a stubborn or lethargic person - to do one thing and not another, so perhaps the brain, when more or less stalled, resorts to self-stimulation by the production or retrieval of information. And the success in both cases, of course, would depend on the relevance and abundance of the information produced. Something like this may well go on, but if it does, awareness of the information, or saying the information to oneself with feeling would add nothing to the process. As suggested in Chapter 8, it may be that any information that is boosted must come to awareness as a matter of physical fact, but then it is still the boosting and not the awareness itself that is necessary. The notion that must be avoided is that awareness is in any way a centre <u>from</u> which efficacious signals, volitions, or any sort of psychic radiation emanates.

Anscombe touches on this, but in quite different terms.

> We can imagine an intention which is a purely interior matter nevertheless changing the whole character of certain things. A contemptuous thought might enter a man's mind so that he meant his polite and affectionate behaviour to someone on a particular occasion

> only ironically, without there being any outward sign of this (for perhaps he did not venture to give any outward sign) ... Let us suppose that the thought in his mind is 'you silly little twit!' Now here too, it is not enough that these words should occur to him. He has to mean them. This shews once more, that you cannot take any performance (even an interior performance) as itself an act of intention ... (p. 48-9.)

There is an internal difference, quite clearly, between just saying "you silly little twit!" and meaning it, but this difference is not itself a performance. The difference, I propose, depends on what the function is of the part of the brain that produces this message in awareness. If the message is produced in the course of the brain's maintaining a particular antagonistic state, if the production of this message is caused by some neural activity that, say, brings into play stored information on the shortcomings of the "silly little twit", then the "thought is meant". If on the other hand the message is produced in the course of, say, mere experimentation, such as seeing if it is in fact possible to be polite while thinking "you silly little twit" and not meaning it, then the "thought is not meant". Suppose the thinker of this phrase has been trying to think of a four-word phrase with internal assonance; suppose in other words someone has just said "give me a four-word phrase" with as much assonance as

"Philip spilt the milk", or for one reason or another this task has just occurred to the person. The neural mechanism that produces the message "you silly little twit", like the neural mechanism that produced in me "Philip spilt the milk", has, or can have, virtually no other effect on behaviour or neural state than the production of the words. The activity involved does not influence or mesh with any other activity.

How does one know whether one means it or not? One knows this simply because one knows what one is about; and one knows this <u>only</u> by knowing what messages preceded and followed the message in question. Imagine a person all of a sudden finding himself saying to himself "you silly little twit". What if no other rancorous thoughts had been going through his head; what if there was no obvious candidate for the epithet; what if further the thinker had not just been aware of thinking he would try this little experiment? Could there be anything intrinsic in the mere unheralded, unaccompanied phrase occurring in his awareness that would tell him whether or not he meant it? Strange isolated thoughts do spring to people's minds occasionally, and they can be totally baffled as to the meaning or importance of these thoughts.

The point that should survive is that awareness is not the home or origin of intentions or volitions. In fact, we have only limited and fallible access to the mechanisms that direct our behaviour. Nothing that goes on in awareness can be construed as an act of will or a volition, and there is not much point in identifying any neural activities outside awareness as volitions. Why burden a fairly clear and straightforward concept with the name of a notoriously foggy and unsatisfactory concept?

44. Voluntary and Intentional.

The class of intentional actions has now been described as the class of actions of which the actor has motor knowledge and for which he is prepared to offer reasons. The question remains whether there is a difference of functional importance between intentional actions and unintentional or involuntary actions. If not, our withholding of praise and blame from unintentional actions is unfounded.

A brief look at the range of bodily motions and actions shows that there is no clear-cut line between the intentional and unintentional or the voluntary and involuntary as far as control or influence is concerned. The most likely candidates for involuntary actions are the working of our innards. Digestive processes, the pumping of blood, and all the various secretions of glands are as remote from our "minds" as anything that occurs in our bodies. The nervous system regulating these activities is called the autonomic nervous system, but of course it is not entirely self-governing. It is connected to the central nervous system, and events in the highest levels of the brain can influence the waxing and waning of these inner processes. Worry, fear, and even such elusive phenomena as optimism and guilt have been pointed to as influences on the autonomic processes. Looking at a picture of food can cause salivation. There is clearly room for argument here if one wants to apply the ordinary word "voluntary" to some of these phenomena. They are not completely outside the province of rational control.

Reflexes such as the knee jerk and blinking are generally called involuntary because we cannot prevent their occurrence. But this is not quite so. The so-called reflexes are controlled by neural circuits that differ from the circuits controlling voluntary motions mainly in complexity and length. The reason one cannot ordinarily prevent a flinch or a blink is because the efferent response to the stimulus is fired before the stimulus can arrive at higher levels of control. One can brace oneself ahead of time and refrain from flinching or blinking.

It might be thought that actions or motions are involuntary if they cannot be prevented after the relevant stimulation has occurred. But this is true of any action. Who is to say that I could have inhibited or prevented the utterance of some retort, once I have made it? We say "You could have kept from laughing - if you had wanted to". To suggest that I could have prevented it is only to suggest that there could have been another neural situation in my brain at the time of stimulation such that the response to stimulation would be different, which is obvious. Whatever state my brain is in determines how I react to the stimulus, and while it is perfectly true that for any response to stimulus my brain could have been in a different state and produced a different response, it is not true that given the state my brain was in, some other response could have been produced.[1] But I do not wish to lean on this point. Even if some acceptable way is found to draw a dividing line between voluntary and involuntary, there remains the division between intentional and unintentional to be examined.

It has been a recurring theme of this thesis that awareness and control are only circumstantially related. A

---

1 For the physicalist position on free will see Scriven, in Hook, op. cit., pp. 121-124, and Smart, op. cit., pp. 120-130.

creature without any power of communication could, conceivably, play the piano like Horowitz, and such a creature would have no motor knowledge or awareness, since he would have no awareness line. The most sophisticated animal behaviour is controlled without the benefit of awareness, and even reasoning was seen not to depend on it. And yet the distinction between intentional and unintentional actions boils down to the fact that we are aware by motor signals of our intentional actions and not of our unintentional actions.

Consider the following case of a voluntary, but not necessarily intentional, action. I walk past the dining room table, see an apple on it, pick it up and start eating it. I may do this in a variety of ways. I may think: "Ah, an apple. It won't spoil my appetite. Besides, an apple a day ... ", and then eat it. Or I may think: "Ah, an apple. Looks delicious. Think I'll eat it", and then eat it. Or I may just "absentmindedly" pick up the apple and eat it while thinking of other things, without saying anything to myself, or even being aware that I am eating an apple. Sooner or later some signal is bound to penetrate to the awareness line, so that I will be aware that I am biting into an apple, but I may get the apple to my mouth and even take a bite before it "dawns on me" what I am doing.

If someone asks me in the first case why I am eating the apple, I can say "because it's good for me, and it won't spoil my appetite, and I like apples", and in this case of giving reasons, I would be actually reporting thoughts that I had been aware of when I picked up the apple. The action would qualify as intentional. In the second case I might answer "Oh, just because it looked so good", and Anscombe would say, I believe, that this was a borderline case of giving reasons; it is also somewhat like a case of giving just "mental causes" (I was just prompted to pick up the apple, and did not reason about it). The action would probably be classed as voluntary, but not really intentional. In the third case, I would say - if the interrogator caught me quickly enough - "Oh, am I eating an apple? So I am.", and the action would fall into the unintentional category.

But in each case the control of the action depended on the same stored knowledge, the same neural operations that I have called reasoning operations. In the third case the visual stimuli from the apple would not have prompted me to pick up the apple and bring it to my mouth if I had learned earlier that apples were bad for me, or tasted terrible. I would not have picked up a raw egg and started to eat it. I would not have been prompted to put a box of matches in my mouth. I would not have picked up the apple in a stranger's home, or in the course of proposing marriage. I did not "consciously reason" that the time was right for eating an apple; but my brain determined that it was. In the first two cases the action was accompanied by a bit of speech centre activity, but the control or direction of the action was not in the speech centre. The only difference between the cases is that in the first two some part of the controls can be reported - fallibly. Perhaps in the first two cases I am controlling my activity with some greater degree of surety because I am giving my brain more time and perhaps more potent signals to deal with the situation, but even then the actual awareness of the thoughts makes no difference.

45. <u>The Importance of Intentional Actions</u>.

The importance of intentional actions is due not to any special controls a person has over his own actions when he is aware of them, but to the controls or influence that <u>another</u> person can have over a person's actions if that person is aware of his actions. The key to intentional action is verbal stimulation directed at other people.

Verbal stimulation contributes to behaviour control in much the same way non-verbal stimulation does. High level motor signals have been likened to orders, but a verbal order, telling someone to do something, does not have the same function at all as the motor signal. It is a bit of information the contribution of which depends on the pre-existing neural organizations and states; the order may be obeyed or disobeyed. In extremely docile or dependent people or in the face of overpowering authority or when one is completely frazzled, the verbal stimulation may in fact contribute so strongly to the firing of motor complexes that it is virtually like pushing a button. In such a case it would be tempting to say the order causes the action, just as the motor order causes the action, rather than that the order as verbal stimulation merely contributed to the cause of the action. But there are always conditions in effect that allow a stimulus to contribute in the way it does; some contributions are just more powerful than others. Verbal stimulation can never be as direct as the tap on the knee that triggers the knee jerk.

In order for verbal stimulation, as for non-verbal stimulation, to contribute to behaviour control, the behaviour in question must be amenable to influence and the stimulation must be relevant. Behaviour that is not acquired or learned behaviour but controlled by hereditary mechanisms, is not very amenable to influence by verbal stimulation. Shivering is a strong example, laughing and crying are more borderline cases. Babies laugh and cry and people in disparate cultures laugh and cry; there is every reason to believe that this is not learned or acquired behaviour, but part of the brain's built-in response mechanisms.

Although there is a difference between crying when the script says "Cry here" and crying when one hears "Smith just died", and between laughing at a funny joke and laughing politely at an unfunny joke, all these are verbally stimulated. We say that in one case of laughing or crying the action is intentional, and in the other case not. Consider the answers to the question "Why are you crying?". If the answer is "because it says 'Cry here' in the script" the crying is intentional, since a reason is given. The answer "because I am sad" leaves the way open for the further question and answer "Why are you sad?", "Because Smith just died". This appears on the surface to be giving reasons for the crying, but there are two non-rational jokers in the chain, revealed by the additional questions "Why do you cry when you're sad?" and "Why should your friend Smith's death make you sad?". There is no verbal stimulation, and little non-verbal stimulation, that can be relied on to influence crying or laughing. No good news or funny repartee will have much effect in stopping crying, and arguing against crying is quite hopeless. Similarly, threats and pleadings have little effect on uncontrollable hilarity. But what about controllable hilarity? That is why laughing and crying are borderline cases, for depending on the intensity of the emotion, the behaviour is amenable to influence by verbal stimulation.

Verbal stimulation must always be relevant; it must strike at the actual controls, and not just blindly. When actions are intentional, one can report some part of the controls, and these reports allow others to aim verbal stimulation with some assurance of accuracy and efficacy.

If the woman in the next room is making a horrendous screeching, and I ask "Why are you doing that?" and she replies "Don't you recognize it? It's '<u>Vissi d'arte</u>'. I'm having my audition next week", I will not attempt to influence her behaviour with such verbal stimulation as "Hadn't you better see a doctor?", or "Stop that terrible screaming". Depending on my subtlety I will say "It's terrible. Have some consideration for others.", or "Would you mind rehearsing in the cellar?", or "Don't you think deep breathing exercises are the best preparation?" or "I think the <u>pianissimo</u> parts are the only parts that need work". If her answer is "An anvil just fell on my foot", I will recognize that verbal stimulation is beside the point and that my best bet is to get some anaesthetic into her foot as soon as possible.

Anscombe says, "Roughly speaking, it establishes something as a reason if one argues against it." (p. 24.) This is the basis of importance of intentional acts; they are acts one can argue against. It does no good to argué against epileptic fits or screaming when in pain - because the controls of this behaviour are quite immune to anything one can say. Yet if one's behaviour devolves on a rational error or a bit of misinformation, arguing against the act can alter or influence the controls - provided the argument is directed at the right controls. The geometry teacher cannot correct a student's miscalculations just on the basis of the incorrect result, but must know the steps the student went through.[1]

Since I can never know my own controls with certainty, I can never know another's controls with certainty. If I am whistling softly and tapping a pencil on a desk and someone says "You're just doing that to annoy me", I may know that no such thought entered my head, and perhaps that some other thought entered my head, such as "This ditty always reminds me of Paris", and I can reply "No, I wasn't trying

1 Cf. Smart, op. cit., p. 122.

to annoy you", but I may be wrong and my suffering listener may be right, although I am on a little better ground than he is. Of course, I was not intentionally annoying my listener, since no signal remotely to that effect crossed the awareness line, but unknown to me the controls of my behaviour may be such that the information that he is annoyed amounts to reinforcing feedback.

Margolis stresses that there is a difference between the ordinary and the "psychoanalytic" notion of intention; the psychoanalyst will often say that one is <u>trying</u> to do something that one honestly believes one is not trying to do.[1] The ordinary use of "intentional", the use explored by Anscombe, does generally allow that actions are not intentional if a person in all honesty claims not to be aware of doing them. The ordinary use, that is, allows the subject veto (see Chapter 3), but only up to a point. The stronger the external evidence that an action is intentional - in the sense that it is being done for a particular purpose - the less likely we are to grant the actor the last word. We say things like "He's been trying to hurt Brown's reputation for years".[2] This amounts to amateur psychoanalysis, and it is from this sort of usage, no doubt, that

1 "Intention, Consciousness, and Action", <u>Methodos</u>, XIV, 1962.
2 <u>Ibid</u>.

psychoanalysis derives its theory of unconscious intentions or motives. The rigorous establishment of this theory must depend on a great deal more neurological knowledge than is now available, for the psychoanalytic use of "intentional" must purport to describe human actions under the descriptions carried by their <u>actual</u> directing signals, and not just under the descriptions carried by the signals that cross the awareness line. Of course even without the neurological backing, the psychoanalytic use can be quite well backed up by other considerations, such as demonstrably good results in therapy.

Anscombe says that "the concept of voluntary or intentional action would not exist, if the question 'Why?', with answers that give reasons for acting, did not." (p. 34.) Or as I have been saying, the concept of intention arises with the activity of determining the controls of behaviour. The <u>ordinary</u> concept of intention, which is dependent on the concepts of awareness and rational controls, is useful because, for the average man, the best hope of contributing to the control of another's behaviour by verbal stimulation lies in aiming his stimulation only at the controls that can easily be altered.

Through the ages the primary method of training and contributing to the control of other people has been fairly straightforward verbal stimulation aimed at the most accessible controls - those of which one can make some report. The primacy of this method has been acknowledged and endorsed in the concept of responsibility. We are only responsible for what is intentional, for what someone in authority, either parents or state, could have told us to do or not to do. As the word suggests, we are only responsible for what we can <u>answer for</u>, and these answers are answers to Anscombe's question "Why?". The exculpation of the insane is founded on such opinions as "It's no use talking to him. He's mad", "He won't listen to reason", "Arguing will get you nowhere", "He didn't know what he was doing; he is not responsible for his acts". Responsibility is founded on the general assessment of the limits of the contributions one can make to another's behaviour control.

It is conceivable that this could change, and perhaps it already is changing. As new ways of effectively directing verbal stimulation and new ways of contributing to behaviour control through non-verbal stimulation and tinkering with the brain are developed, the concept of responsibility may dissolve or shift with changes in the concept of intentional action.

CHAPTER 12.

Knowing.

46. Knowing and Truth.

The characterizing question of epistemology is: What is knowledge? Attempts to answer this vague and general question have often led philosophers into the field of philosophical psychology, as their question became transmuted imperceptibly into the quite different questions: How do we know what we know? How do we perceive what we perceive? What is perception of? What is consciousness? What are sensations? Once the epistemological question has taken this turn toward psychology, the answerer's ignorance of psychology and physiology becomes indefensible, as I argued in Chapter 1. The epistemologist need not take this turn, but the temptation is almost irresistible, largely because of a fundamental inconsistency in the ordinary concept of knowing.

The ordinary use of "know" carries with it the claim that what is known is true. If I claim to know that p and "p" turns out to be false, my claim of knowledge is disallowed. It will be said that I only believed that p, but did not know that p. Yet at the same time one supposes that the class of things known by a person must be in a special psychological position, so that at least a person can tell which things he knows and which things he only believes, for example. We say things like "to the best of my knowledge", "It is safer to act on knowledge than on mere belief", and most importantly, "What do you know about p?". One assumes, in speaking in these ways, that one can tell what one knows, and that what one knows is of particular functional importance to one's behaviour.

These two notions about knowledge, the truth condition and the ability of the individual to tell knowledge from belief, are incompatible. The actual truth or falsity of information stored or incoming in a brain cannot have any effect on the functioning of the brain. Truth is not a functional characteristic of neural signals or storage patterns; signals do what they do whether or not they are veridical. The brain may test the relative veracity of signals, by rejecting one or another of two incompatible signals, but those signals that "pass the tests" are only apparently true. What matters in the brain is the functional status of information, what the information is in a position to direct or affect; information has the status it has regardless of whether it deserves its status. Similarly, a factory may test and retest its products in an effort to ensure that all goods sold meet certain specifications; but the items that have the status of having passed the tests (say, all the items in the "ready for market" warehouse) are only apparently up to specifications. The functional status of these goods is the status of being salable or of being the "finished product", but not of being the perfect product.

There is no reason why the functional status of information stored in the brain cannot admit of degrees. Intuitively, we feel that there are degrees of belief; we are surer of some things than of others. A man can order a group of statements according to how willing he would be to make his life or career or future happiness depend on their truth. There is nothing to suggest that it would be physically impossible for a person to determine classes of things of varying degrees of surety for him, for this is the sort of difference that could have a functional distinction in the brain behind it. But a man simply could not "intuit" or "introspect" the class of things not only believed but also actually true.

Hence when called upon to produce one's knowledge, one can do no better than to produce what one believes to be true - what one would act on - and whether or not what one believes in this way to be true <u>is</u> true does not affect its belonging to the category of things one will produce as knowledge when asked. Even the denial that follows if what a person says is deemed false does not show that he erred in saying it. As Ayer says, "The discovery of the error refutes the claim to knowledge, but it does not prove that the claim was not, in the circumstances, legitimately made."[1] What Ayer has described is not merely a curiosity of ordinary language, but an outright inconsistency that must be remedied.

There is an option then: either what is known must be true so that the class of things known by a person cannot be determined by him, in which case it is absurd to ask him what he knows; or the truth condition must go, allowing a person to say he knows whatever he places a high degree of commitment on, whatever he would stake good money on. There is a temptation to say that the latter characterization of knowing is the psychological characterization; it sets aside a class of stored information of particular functional importance to the person. Then the former characterization might be called the epistemological characterization, and what it sets aside is the body of truth believed in concert by the peoples of the world. This epistemological characterization is of dubious value, except perhaps as a myth, since who is to determine the

---

1 The Problem of Knowledge, 1956, p. 43.

extent of this body of truth, if each man cannot determine what he himself knows?

The difficulty with the truth condition does not end here, for since the truth condition is a condition of the ordinary concept of knowing, we are all cognizant of it, and hence, when asked to say what we know, we do our best to say only what is true, and not just what we believe to be true. But this is impossible. If one is asked "Which of the things you believe are actually true?" one can only answer that to the best of one's knowledge _all_ of them are true. One does not believe things one believes to be false.

What then is a person doing when he claims to know some things, but only believe others? A person can order the things he believes according to the strength of his belief in them, but where on the scale is the line to be drawn marking off belief from knowledge? Clearly, no matter how high on the scale the line is drawn, the information above the line is not ensured of _being_ true, but only of being very, very strongly believed to be true. Ryle observes that "know" is often taken to be an achievement verb, like "catch" or "find"; one does not say that one knows incorrectly or unsuccessfully. The connotations of "knowing" is often that of _finding out_.[1] In fact, the only achievement

---

1 _Op. cit._, pp. 149-52.

there can be here is the production of beliefs that are satisfactory to oneself. The achievement of knowing can only be the achievement of apparent agreement with the authorities or the community at large. One draws the line between knowledge and belief pragmatically, hoping that everything above the line not only conforms with whatever one's interlocutors believe, but also that no evidence will come up to refute both oneself and one's fellow men.

One may hope to clarify this situation by appealing to another traditional condition of knowledge: the condition of adequate evidence. I may believe that I am a direct descendant of Charlemagne, and even if this were true, I do not know it, since I do not have adequate evidence - or any evidence at all. The condition of adequate evidence is certainly not a condition that has been discovered empirically to govern our claims to knowledge. At the very highest point on my scale of beliefs are many facts for which I have not the slightest evidence, and yet I quite properly claim to know them. I cannot remember having read of them, or having heard of them from reputable authority, or having seen at first hand any evidence at all.[1] And even if I have read them in newspapers, or been told by my parents, and can remember the specific event of having

1 Cf. Putnam, op. cit., p. 156.

found out, does this constitute adequate evidence? One might emend the condition so that a person claiming to know need not himself have the evidence, but there must be adequate evidence. Then, however, it is perhaps the case that there is, unknown to me, adequate evidence that I am a direct descendant of Charlemagne - and one does not want to say I know this. Perhaps I must know that someone has adequate evidence, but in fact I often do not know this, and furthermore, if my knowledge of evidence is included in this condition for knowledge, a regress sets in.

Clearly, the condition of adequate evidence is normative rather than descriptive. One should have adequate evidence for whatever one claims to know. But of course, one can only believe that one has adequate evidence, so the stipulation of the condition is not the promulgation of a law to be followed, but an exhortation. Until standards are set down for adequate evidence, the rule that one should claim to know only what one has adequate evidence for has no application, and once standards are set down, one can only do one's best to meet the standards in each instance. The normative rule can only be an admonition for prudence: look before you leap. There are other difficulties with this condition but they are strictly epistemological, and can be considered without particular reference to the physical situation or activity of knowers, so they will not be discussed here. One can ask, for example, just what constitutes adequacy of evidence, or to what end is this adequacy relative.

One might object to this view, the view that one cannot intuit a distinction between the true and the believed to be true, that some of the things we believe have their certainty built-in, and hence can be distinguished as true and therefore as known, not merely believed. These candidates for knowledge are known variously as _a priori_ truths, tautologies, logical truths or analytic truths. But they do not qualify, and this is alluded to quite directly by the word "analytic"; one must analyze these facts, and one's analysis can be mistaken. As Ayer points out, "From the fact that _a priori_ statements, if they are true, are unassailable [in that if true, there are no circumstances in which they could be false], it does not follow that they are immune from doubt. For, as we have already remarked, it is possible to make mistakes in mathematics or in logic. It is possible to believe an _a priori_ statement to be true when it is not."[1] Our knowledge that p is certain cannot be certain.

Another class of putative certainties is the class of

---

1 _Ibid._, p. 42.

things of which one is aware. In Chapter 7 it was shown in just what way one could be infallible in the expression of the contents of awareness. The infallibility disappears, however, as soon as the moment is past; memories of what one was aware of are not infallible. Furthermore, in putting the contents of awareness to any use, one must accept the arriving signals as veridical, which brings in the possibility of error again. The moment one starts using these infallible utterances or "thoughts" as building blocks of knowledge, their infallibility disappears.

The futility of these traditional conditions for knowledge can be revealed by a cautionary tale. A group of men are being taken in by the old shell game. Andrew watches the play for a while, and then says "I know where the pea is. It's under the middle shell". He places his money on the middle shell. Bill says "The evidence of my eyes in such matters is not adequate evidence for claiming to <u>know</u> where the pea is, but I do believe it is under the middle shell, and am willing to stake my money on it". Charlie is a philosophical sceptic; he says "I do not <u>know</u> that there is even a shell game going on. I only know there <u>seems to be</u> a shell game going on. I do, however, <u>believe</u> that my senses are veridical, there is a game going on, and the pea is under the middle shell, so I will seem to place the money I seem to have on it". David says "I am perfectly willing to say I know the pea is under the middle shell - I am just as sure of it as I am that I had eggs for breakfast, but I'm very cautious about betting, and will not bet in this situation". The bettors lose, of course.

Andrew acted on what he called knowledge, and lost. Bill refused to claim knowledge, but lost. Charlie lost, and there will be small solace for him in his philosophical conviction that he does not <u>know</u> he has been swindled. David comes out the winner, although he was willing to claim knowledge. Suppose Andrew says "That's strange. I could have sworn that I <u>knew</u> where the pea was, but I was wrong, so I must have misidentified my experience as being an experience of knowing rather than believing". This is surely nonsense; identification and misidentification just do not apply here. Suppose Bill says "Next time I'll be careful not to put my money on something I don't <u>know</u>, however strongly I believe it". He will be no more able next time to identify his knowledge than he was this time; he may be less gullible next time, but this will not involve making any identifications of knowledge. Since Charlie's standards for knowledge are so exclusive, he must get through life by acting on "mere beliefs", and hence he does not <u>use</u>

whatever standards he claims to have set up. David may seem at first to have sloppy standards, but perhaps he realizes that he can do no better than he does at present. He recognizes that there is no difference for him between what he knows and what he believes he knows; he does not search for truth by first searching for what he knows and then arguing that since he knows it, it must be true.

The epistemological dilemmas about truth and evidence are not to be imported into psychological or physiological attempts to describe the functional status of information stored in the brain, for truth and evidence are not relevant to the questions concerning this status. Nor will the epistemologist gain anything by moving his questions about truth and evidence into the psychological sphere, for conversely, what status information has is not relevant to the questions of truth and evidence. The epistemological questions are quite independent of the psychological questions, and can be answered independently, provided that cognizance is taken of one central psychological fact: what one knows, one knows only "to the best of one's knowledge".

In the rest of the chapter, I shall discuss the information stored in a person's brain in such a position or with such a status that the person will act on the basis of it. I shall call this information <u>known</u> by the person.

I say "known" rather than merely "believed" because I wish to exclude from consideration the information at the bottom of the scale of beliefs, the frivolous beliefs (about which horse will win the Gold Cup, whether the coin will fall heads or tails) on which we will base <u>some</u> acts, but not very many. The extension of this class of things known is flexible on the scale, and I wish merely to stress the upper end of the scale.

It might be objected that this characterization of what is known does not allow for people who tell a lie and then support it by acting always on the basis of it, as if they believed it. How does the unbelieved information differ in its functional status from believed information? The error in this objection is in assuming that we store the lies we may act on. One does not and cannot store a lie or any information one holds to be false. One stores what is believed, and then when one lies or lives falsely, one acts on the basis of the believed information, and the form one's acts take is the form of dissembling acts. I may store the information that I lied yesterday in saying that p, and then today I will behave on the basis of this knowledge by supporting my lie of yesterday. If the lie is said to be stored at all, what is stored is the believed information that it is a lie, and I told it.

47. <u>Understanding</u>.

Storage of information does not by itself constitute knowing; dictionaries and encyclopedias and libraries can be said to be stores of information, but they do not know the information stored. The condition that will restrict knowledge in the desired way is understanding. One must understand what one knows. In §34 a distinction was made between what might be called mere verbal conditioning and true understanding. The computer programmed to translate from English into Russian does not understand its input, because it does not have any reaction to "I just murdered my uncle". Even if it were programmed to produce paraphrases, like "You have recently slain the brother of one of your parents", it would not be said to understand. If, however, the computer immediately made a discreet telephone call to police headquarters, one would be tempted to say it had understood the sentence in quite the fullest sense, but only if it also had the capacity to do other quite different things with different input. If it is merely the local ADIAC computer (Apparatus of Dubious Intelligence for Acknowledging Confessions), no one will grant it understanding.

Only a being that is non-verbally active in the world could meet our intuitive requirements for understanding.

47. Understanding.

Storage of information does not by itself constitute knowing; dictionaries and encyclopedias and libraries can be said to be stores of information, but they do not know the information stored. The condition that will restrict knowledge in the desired way is understanding. One must understand what one knows. In §34 a distinction was made between what might be called mere verbal conditioning and true understanding. The computer programmed to translate from English into Russian does not understand its input, because it does not have any reaction to "I just murdered my uncle". Even if it were programmed to produce paraphrases, like "You have recently slain the brother of one of your parents", it would not be said to understand. If, however, the computer immediately made a discreet telephone call to police headquarters, one would be tempted to say it had understood the sentence in quite the fullest sense, but only if it also had the capacity to do other quite different things with different input. If it is merely the local ADIAC computer (Apparatus of Dubious Intelligence for Acknowledging Confessions), no one will grant it understanding.

Only a being that is non-verbally active in the world could meet our intuitive requirements for understanding.

The links between verbal and non-verbal behaviour and stimulation are particularly evident in this case. One feels that a computer whose only input and output was written on tape would always be blind to the meaning of the input; it might grasp all the verbal connections, but it would have no information on the associations between words and things. It might be able to paraphrase descriptions of the Taj Mahal, or might produce an output like "The Taj Mahal must be beautiful", but one wants the computer also to produce outputs like "Take me there; I want to see for myself", and such outputs would be a hoax if the computer did not have some perceptual apparatus.

It may seem that understanding must require a particularly sublime capacity for wedding words and deeds and things, a capacity so high that no machine, save a man-made machine, could ever be said to understand. How is the machine to understand the word "love" or "beauty" if it cannot experience these? But must the machine understand <u>all</u> words in order to understand? There are many men who have never experienced love or beauty, and perhaps one wants to say their understanding of these words is defective, but surely they understand many other words. Few people understand the pronouncements of nuclear physics, but this does not bar them from understanding in general.

There are some strictly behavioural minimum conditions for understanding. A being cannot be said to understand some jot of information if it cannot behave in a variety of appropriate ways on receipt of this information, and the being still cannot be said to understand this information unless it also can behave in a variety of appropriate ways on receipt of other information. One could not build, a machine with the single function of understanding one particular sentence. These conditions work as well for human beings, where the question is not whether the being has any capacity for understanding, but whether some particular information is understood, If Jones says "Smith is here", no one will allow that Jones understands and hence knows that Smith is here unless Jones can also say and do a variety of other things with his knowledge. He must be able to assert, for instance, "Smith is not in Siam", "Smith is that friend of Black's", or "By 'here' I mean 'in town', not 'in this room'". If Jones knows Smith is here he must be able to point him out, or at least direct the search party. If no such corroborating behaviour is in the offing, Jones may be no more than a parrot or a phonograph. Jones may announce or assent to the statement "There's a mad dog in the next room", but if he then strolls into the next room without taking any precautions, it is not at all clear that he knew or understood what he appeared to know and understand.

The particular tests that must be passed in any one case are not strictly determined by the information in question. What tests must be passed depend largely on what else the person knows and understands. And whereas a great deal of the corroborating behaviour can be verbal behaviour - explaining, asserting related statements, paraphrasing, and expanding on the subject - if there are available non-verbal tests and they are failed, the verbal testimony will be shaken. Much information, of course, is so intimately verbal, in being information about verbal states of affairs, that no strictly non-verbal behaviour could tend to corroborate the claim to knowledge: e.g., the information that yesterday was (called) Friday, this place is (called) Oxford, a man called Wren built the building called the Sheldonian.

It is difficult to decide whether these necessary behavioural conditions should be considered sufficient conditions as well. It was suggested earlier that understanding admits of degrees. What are the conditions that suffice to show that a child understands his own utterance: "Daddy is a doctor"? Must the child produce paraphrases, or expand on the subject by saying his father cures sick people? Or is it enough if the child knows that Daddy's being a doctor precludes his being a butcher, a baker, a candlestick maker? Surely the child's understanding of what it is to be a doctor (and what it is to be a father) will grow through the years, and hence his understanding of the sentence will grow; he will be able to offer more and more related behaviour, verbal and otherwise. If there can be degrees of understanding, then no sufficient conditions can be set down for it unless an arbitrary line of some kind is drawn. Nor can any strict group of necessary conditions be set down.

48. The Content of Knowledge.

If understanding admits of degrees, and if understanding is a condition of knowing, then knowing must also admit of degrees, as was suggested in §31. This presents a number of difficulties for the ordinary concept of knowing. I have been equivocating in this chapter between the idea of information or facts as statements and the idea of information in the sense of functional potentials for the direction of behaviour. I have talked very vaguely about things known. This must now be cleared up.

What is the relationship between understanding a sentence and having stored information? More specifically, what is the fact that is known when a person asserts, understands, and believes to be true, some sentence? If one can talk of a fact at all, the fact known when one asserts a sentence must vary from speaker to speaker, depending on the degree of understanding. This bodes ill for facts, but there is worse to come. If a sentence can express a different fact for different people, the facts stored by the vehicle of sentences in encyclopedias must vary from reader to reader. If some means is found for stabilizing "facts" in a book, then these facts will not be what is known by a person.

A child with the rudiments of arithmetic knows just a little about the number four, but the effect of this small knowledge is that the child can reel off reports of his knowledge ad nauseam: "four is half of eight, ... four is 1/250,000 of a million". None of these reports, and no finite collection of them, exhausts his knowledge about the number four, and since he is not an advanced student of mathematics we cannot expect him to offer, assent to, or even understand statements about the real number system, the infinite number of multiples of four - in short the sort of statements that might be held to generalize and exhaust his knowledge. What fact or facts can we say the child knows? Does he know an infinite number of facts? Or just one or two rather general facts? If the latter, he knows facts the expressions of which in his native tongue he probably does not even understand.

An encyclopedia - a very small one - might be held to store just one fact (if it consisted of one printed sentence), but a person could not be held to know just one fact. The knowledge of one fact could not exist by itself because the fact could not be used, and hence could not be understood. What a person can use his stored information for depends on what other stored information he has, what else he knows. The things we do with our knowledge are quite discrete, but our knowledge itself does not divide into neat, independent parts.

If there is a concept of facts to be salvaged for some uses, it will not have any handy theoretical application to human knowledge, to what is stored by a person in his brain. For the pragmatic purpose of testing knowledge for university entrance, for example, it may be that writing or assenting to statements (as in true-or-false tests) can be seen as a measure of what facts are known, but such a method simply ignores the factors of comprehension, ability to use the information, and lucky guesses. The metaphor of the walking encyclopedia is simply not to be trusted, however facts may be anchored for encyclopedias.

Philosophers have often attempted to tie facts to particular placements of things in the universe (the cat on the mat), particular time-slices of spatio-temporal reality, particular concatenations of qualities presented to the senses, but all these attempts to tie information to states of affairs have been doomed to failure for the simple reason that for any state of affairs or placement of things, in the universe, there are many very different things that can be said to describe this situation. One supposes, for example, that there is a particular scene or time-slice that contains the stimulus conditions for the sentence "Smith, Jones' uncle, has died", and that if this scene is properly limited, it determines the information carried by that sentence. One imagines the stopping of a heart in a tension-filled room, and hopes somehow to ostend, if not describe in words, just those features of the situation that determine a particular jot of information, the information carried by the sentence. But however these features are set out, they will also be the stimulus conditions for many other sentences, such as "The shock treatment has failed", "The local Rotary Club is without a treasurer", "The heart of the only reclining person in room 235 has ceased to beat", and so forth. Now if the information carried by a sentence were wholly determined by these stimulus conditions, then any sentence with stimulus conditions x could be expected to transmit the information x' one wanted to transmit. But if someone tells Jones that the local Rotary Club is now without a treasurer, and if Jones has no idea his uncle was treasurer, he will not have been informed of his uncle's death, and this is what is wanted.

These attempts to tie information strictly to states of affairs fail because the function of the intermediaries between sentences and states of affairs, the message-makers and receivers, is not taken into consideration. A message picks out some feature of the state of affairs that is functionally important for some receiving system. Similarly, a neural impulse acquires no content simply by occurring if and only if certain stimulus conditions are in effect; it is only a signal when it goes on to contribute to other functions. The freezing of a pond is not in itself a signal to the effect that the temperature of the water is below the freezing point, since if there is no receiving system present, the freezing is not, and does not contribute to, a functional change in any receiving system.

The new field of information theory, while it does not provide exactly the concepts and methods needed to describe human knowing, does suggest several important points. The theorems of information theory concern the reliability of transmission systems, and in order to provide a quantitative measure of reliability, a method has been developed for measuring amounts of information. Significantly, the amount of information is not directly a function of stimulus conditions or of, say, the internal semantic structure of the signals (which are treated holistically, and are usually, in any case, "offs" and "ons" in a binary system), but of the degree of uncertainty diminished in the receiver. The receiver is given the task of singling out some individual or individuals from a limited ensemble or class of possibilities, e.g., finding out what day of the week it is. The signals received serve to exclude possibilities (e.g., the signal "It is not a weekday"), thus reducing the ensemble, or one signal can single out the individual, solving the problem in one step.

> How much information is in the statement 'This is Friday'? We now know that we must first determine the context. Suppose our ensemble was 'The days falling between Thursday and Saturday'. Such an ensemble has one member, so that
>
> $I = \log_2 1$ bits [the unit of information]
>
> $= 0$ bits.
>
> The statement, then, contains no information.
>
> In another context the result could be different. Suppose we know that since we are working it is neither Saturday nor Sunday. In this case, our ensemble has five equiprobable members, and
>
> $I = \log_2 5$ bits
>
> $= 2.32$ bits.
>
> Finally, let us suppose a man awakens from a coma. He has no idea how long he has been

unconscious, and asks 'What day is it?' The seven possible outcomes are equiprobable, and

$$I = \log_2 7 \text{ bits}$$
$$= 2.81 \text{ bits.}$$ [1]

A signal or message, then, like "This is Friday", informs only relative to its function in ordering an ensemble, and the ensemble is determined by the receiver. To the man who knows that yesterday was Thursday and tomorrow is Saturday, "This is Friday" is no news. This way of determining the amount of information works only for ensembles with a known number of equiprobable members. Thus it is of no use in determining the information content of "Your uncle just died" or most of the sentence tokens occurring in everyday life. In the case of "Your uncle just died" the ensemble might be held to consist of two members, uncle dead or alive, and if a person were waiting for news on the state of his uncle, then to say that the sentence carries $\log_2 2$ or 1 bit of information would make some meagre sense. But in human beings, as opposed to devices with one limited job to do, the receipt of information allows a great many different ensembles to be partially ordered, depending on the knowledge held by the receiver. Thus what seems intuitively true is given some mathematical backing: the information received by human beings when they are spoken

1 Edwards, Information Transmission, 1964, p. 39.

to depends on what they already know, and is not the sort of thing that can be measured in precise amounts. What information is carried by "There is a bottle of good sherry in the kitchen"? Until context is known, and until the knowledge and "purposes" of the receiver are known, there is no answer, and no ordinary context would allow the fixing of plausible finite ensembles for the receiver.

Another point of importance that emerges from this treatment of information is that information need not be carried or stored by sentences made up of semantic elements. Any system of events or states can transmit and store information, and the diversity of information that can be handled depends on the number of different events and states that are possible in the system. An obvious fact, but one that is easily overlooked, is the fact that the capacity of language to store and transmit information (in books in libraries, in speeches and hoardings) is dependent on the existence of non-linguistic means of storing and transmitting information. Information is not preserved in a sentence the way a fossil is preserved in a rock. A closer analogy is with a bank cheque, which does not preserve gold, but can be used to transmit wealth. A bank cheque is only valuable within a system of use, only when all parties agree on its function. A sentence carries information only because it is part of a system that could not exist unless information could be transmitted and stored (in the individual) non-verbally. If information were stored in the brain verbally - by some sort of cerebral tape recorders or printing machines - it would still have to be translated into non-verbal information before it could be of any use. The brain would have to have an understanding-reader-device. Without such a device, words stored in the brain would be as much use as saying "giddyap" to an automobile.[1]

In Chapter 6, the world of content was described as a world of language and its superimposition on neural function was seen as artificial and heuristic, but the world of language is not itself independent. Far from being entirely alien to systems of neural function, the system of content in language is dependent on them for its existence. The means of explanation have in effect come full circle. Content was explained in earlier chapters by reference to messages in language; now messages in language are explained as having content only in virtue of the dependence of language on the functioning of neural systems. This

---

1 Research in the operation of memory in the brain has been relatively fruitless; if, as sometimes seems to be the case, the researchers are looking for some sort of imprinted, coded information, it is no wonder they have been unsuccessful.

circularity is not vicious, but merely serves to point up the fact that the world of content is entirely a world of interpretation, and has no entities above and beyond physical objects, in particular, the appropriately functioning control systems that produce the world of content as an aid to describing their own functions and behaviour.[1]

In spite of this dependence of language on the existence of non-verbal information processing systems, no natural or even quasi-natural language could be "mapped on" to a human neural system systematically. The use of concatenations of semantic elements (sentences, in other words) for the ascription of content to neural signals can never be more than a heuristic procedure of approximation. For ascription to be truly systematic one would have to be able in theory at least to make predictions within the domain of content alone. Given a body of ascriptions of content for some signals, one would have to be able to determine what would <u>follow</u> from these for the brain. But what follows from a sentence in natural language need not at all follow in the brain, for a number of reasons.

The ascription of content to neural signals is

---

1 Cf. Feigl, <u>op. cit.</u>, pp. 417-8, on "aboutness". He seems to fail to realize that semantics must eventually come home to roost in the concepts of control systems and their functions.

illuminating largely because there is an analogy, however imperfect, between the scope of versatility of neural signals and the scope of paraphrase of ordinary sentences. The brain's "understanding" of a neural signal is a matter of utilizing the signal to order and direct a wide variety of behaviour. A person's understanding of a spoken sentence is at least in part a matter of what other messages he can derive from it with what he knows already. The scope of paraphrase for a sentence, however, is indefinitely large, depending on the information one can discover about the words and the things referred to by the words. Thus "Smith has just died" yields "The man, of the family named Smith, who lived at 10 Green Street, no longer has a beating heart", "Smith can no longer perform any of the activities of life", "Jones' uncle is deceased" and so forth. In addition to the sentences that might remotely be called paraphrases, many other things may follow, such as "The Rotary Club has lost its treasurer", and "The heart massage in Surgery B has failed".

If one were to suppose that a person had total knowledge of all relevant facts about Smith and total understanding of the sentence, then one might expect that for any fact that can be shown tonfollow from the sentence (taken as true), it must follow for that person. But no person has total knowledge and understanding, and hence one cannot rely on any particular extrapolations from the original sentence. The same holds true for the brain. Just because a neural signal has one or two functions which suggest ascribing to it a particular sentence as content, it does not follow that everything that can be extrapolated from that sentence follows or can follow for that brain.

The scope of paraphrase allowed for the message would have to be restricted to just those paraphrases for which there was functional backing. The point of restricting scope, however, would be to regulate the prediction of the content of neural continuations, but in order to restrict scope one always would have to look at what the actual neural continuations were. The traffic between the domain of physical function and the domain of content must thus be all in one direction. One cannot argue from content to function (or from content to following content), for the content is always subject to revision or diminution depending on what the functions turn out to be. Furthermore, the ascription of content always assumes the apparent appropriateness of the functions that occur; it leaves no room for inappropriateness, and hence could not be used to predict or explain the occurrence of inappropriate functions.

The functional capacities of a spoken sentence depend to some degree on the words that make it up, and how these words are defined. Just the opposite can be said of neural messages. They have the function they have, whatever content in words one wishes to ascribe to them. Neural signals are not composed of parts that are even remotely semantic parts, and hence the analysis of the semantic parts of ascribed verbal messages does not parallel or replace any analysis of the signals themselves. Having ascribed a certain content to a neural signal, nothing follows from this ascription, whereas having spoken certain words as a message, what follows is to some degree dependent on the words spoken. Hence the hair-splitter who demands absolute rigour in the ascription of content, who insists that the verbal message ascribed should contain just the information carried by the neural signal, is asking for the impossible. There will always be many possible extrapolations from a verbal message, however nicely it is worded, and which of these are accounted for by differentiations in the neural system cannot be stipulated in the verbal message.

49. Saying What I Know.

If what is known is stored (or presently incoming) information, then one would expect that saying what one knows would be somehow transmitting by means of language some of this information. The intuitive account of saying what I know is that I attempt to transmit or impart or share with my listener something somehow held by me as known; I try to produce in another person something (knowledge of something) that I have. A tentative account in terms of information storage and content would be: when I say what I know, I utter a systematic series of sounds that stimulates in the listener (if he uses the same system of sounds and does not reject the input) the production of an information storage pattern similar in content to some storage pattern in myself.

The weasel word in this account is "similar". Since content is to be determined not just by stimulus conditions but also by potentiality (capacity for further function), it would be very rare for the listener to acquire stored information of exactly the same content as that stored in me, since there will always be differences in potentiality unless the listener has an information store which already duplicates mine in every relevant respect save just what I am transmitting. This is not a difficulty peculiar to this way of talking, but a phenomenon that is perfectly apparent however one talks. If I do not know that Tully is Cicero, and say "Cicero denounced Catiline", my listener, if he knows

Tully is Cicero, will in effect come to know more than I was endeavouring to tell him. Should one say that he received more information than I sent? It is plain that his receipt of the message I sent allows him to do things I cannot do, in particular, to store the additional information that Tully denounced Catiline. Such dividends of information do not always hinge on synonymy or identity of reference, as in the Tully-Cicero case. On hearing "Your uncle just died", Jones may be able to infer that he will soon be a rich man, that a certain Mrs. Smith is now a widow, or that a certain surgical operation failed, and however information is to be construed, these dividends will not be equivalent in any way to the information imparted to Jones. Being able to infer these things, Jones will be able to store them, and know them.

What, then, is the similarity of information produced in the receiver when a person says what he knows? At first glance it seems too strong to say that the information produced in the hearer has to duplicate the information stored in the speaker with respect both to possible stimulus conditions and to potentiality of further function. What a person can do with the information seems irrelevant. But in fact it is required that speaker and hearer share relevant information for successful information transmission to occur. A sentence is intended to have a certain approximate effect, which it will have only if the receiver or hearer has certain information. This information that must be held is just the information shared by speaker and hearer that is relevant to the information transmitted.

There is more to this than the fact that if I attempt to communicate in English with a person who speaks no English, I will not succeed in producing in him information similar to information stored in me. Even if the hearer is English, he must also have much the same background of information on the subject of discussion as I have. The sentence "I've found a solution to the problem of other minds", which contains no words that the average adult English speaker would not know, is still unlikely to be informative to a person who does not share with the speaker a background of knowledge of the problem, the speaker's activities, and what might be held to be a solution to the problem. The similarity of information stored need not be complete, for information does not come in a fixed amount, all of which must be transmitted and received. This fact is evident from such common statements as "I understand you, but what you say doesn't tell me very much".

If saying what one knows is attempting to produce in another person an information storage pattern of similar content to one held in one's own brain, there is the question of how such a sophisticated operation could become established in the human repertoire. The establishment and survival of programmes for controlling behaviour have been seen to depend on the "demonstrated" appropriateness of the behaviour to conditions in the stimulus environment. Now if one type of verbal behaviour that arose among the cavemen with the advent of language happened to have the effect of stimulating in others the storage of information similar in content to stored information in one's own brain, the stimulus results of this behaviour would be felicitous, and hence the type of behaviour would be reinforced. That is, the utility of sharing information with one's companions would have numerous manifestations, such as teamwork and versatility of the group, leading to fuller stomachs, greater security and more leisure. Of course verbal behaviour would be unlikely to <u>happen</u> to effect this unless there developed simultaneously some agreement over the use and reference of words. Without conformity of verbal behaviour, few felicitous results would occur.

The adoption of this behaviour would in no way depend on the cavemen's knowledge of the physical mechanisms in the brain that accounted for the appropriateness of the behaviour. Appropriateness must be impressed on the brain, but the theory behind it need not be demonstrated to anyone. The propitious manifestations, which would vary with the degree of success of storage duplication, would by themselves, without the need of an explaining theory, suffice to reinforce the best transmissions and extinguish inappropriate verbal behaviour. At the same time, of course, other stimuli would be coinciding with verbal behaviour to reinforce the appropriateness under certain conditions of other types of verbal behaviour - such as lying. Just as wolves need no theory about the conversion of food into energy in order to develop the appropriate behaviour of food-hunting, man could develop the habit of saying certain things when suitably stimulated which happened to be useful because, unknown to him, they served to transmit neurally stored information. Teaching the appropriate behaviour to each generation of offspring would be an inseparable part of teaching them language. The "common sense" description of what is going on when one says what he knows (e.g., one is communicating ideas held in the mind and believed) could be a much later bit of theorizing.

An important feature of this behaviour of saying what one knows is that its characterization - what sets it apart from some other types of verbal behaviour - involves a description of internal, neural conditions. Whereas the differences between asking questions, giving orders and making statements can be described in terms of external characteristics of grammar, inflection and so forth, the difference between saying what one knows and lying, for example, cannot be described in this way.

Since this characterization is internal (though not <u>necessarily</u> private), a person who tries to prompt another to say what he knows, and who has no access to the internal situation, cannot know that he is getting the desired response to his prompting. But the fact that he does not know the internal characterization of the type of response he desires does not prevent him from desiring the type of response he desires. He may have some vague theory about its characterization, and may say, as part of his prompting, "I want you to share with me the true ideas you have in your soul about X", but what he is really after is the response that has the particular utility that the transmission of neurally stored information would have. Saying what one knows is characterised in common sense by what it is good for, and what it is good for is what the transmission of neurally stored information is good for.

This pragmatic characterization goes some way toward explaining the inclusion of the truth-dependency condition in the ordinary use of "know". The internally characterized behaviour of saying what one knows would not in fact be useful or appropriate, and hence would not survive, unless most reports of knowledge were not only believed to be true, but were true. Thus the psychological conditions for knowing, which I urged in §46 should be distinguished from the epistemological truth condition, are not in fact strictly divorceable from the question of truth. In general and over the long run the truth condition must be met by reports of knowledge if the type of behaviour is to survive. It is only in each particular case that the truth requirement is in principle unfulfillable.

Still the notion of the truth requirement must suffer qualification. This dependence for survival on truth is not dependence on absolute truth, whatever that is, but just on truth relative to the ordinary efficacious stimulation in the environment. As long as the behaviour is in harmony with the stimulation that happens to contribute to neural activity, it will survive, whether or not "the world is as it seems to be". The truths on which the behaviour is dependent are the everyday truths about what we had for dinner last night and where the nearest bus stop is, not the truths of cosmologists and ontologists.

50. Knowing That I Know.

If an observer of behaviour that is putatively an example of saying what one knows cannot, on the basis of external behaviour, know that the behaviour is genuine, can the speaker himself know this? This is a variation on the old question, If I know that p, must I or do I know that I know that p? There are several ways of taking the old question.

There is a difference between having a penny in one's pocket and knowing one has a penny in one's pocket. Then, is there a difference between knowing that p (having the information that p) and knowing that one knows that p? The difference in the case of the penny in the pocket has behavioural manifestations; if one knows one has the penny, there are things one can do that one cannot do otherwise - one can retrieve the penny on demand, for example. But the ability to retrieve stored information has already been included in the definition of "storage". Information that cannot be retrieved and used is not stored, is not known. The idea of having something one cannot get at is of dubious value anyway. If chests of pirate treasure are buried on my property I know not where, do I have them? If I do not even know they are somewhere on my property, am I a rich man? Ordinary usage, which is unclear about "have" in this respect, follows the hard line with "know": if you cannot produce it, you do not know it. My analysis of knowing follows this usage. So in this sense of the question, Must I know that I know that p?, the answer is yes, by definition.

The question might also be taken to ask if, when I know that p, I know <u>for sure</u> that p. One might suppose that knowing one knows that P is a state of greater certainty than just knowing that p (especially if "know" is taken to demand the truth-dependency condition). However, the claim "not only do I know that p, but I know that I know that p" can only be taken to be the claim to have checked and re-checked p, and amounts at most to "My claim to know that p is not to be taken lightly". One does not approach absolute certainty by annexing "I know that I know that ... " to knowledge claims any more than one gets richer by listening to the ring of a single coin over and over again.

To return to the original form of the question, when I say what I know, do I know myself <u>what</u> I am doing (that I am not telling a lie, for example)? First, it is clear that I do not know much if anything about what neural events might be going on in my brain, certainly not enough at any rate to characterize my behaviour as saying what I know. In the future I might be able to know this, if neurology carries us that far, but then my listener might know the same thing; this is not the right sort of answer.

The very occurrence and survival of such behaviour as saying what I know depends on the brain's "knowing" what it is doing. The reinforcement of appropriate behaviour involves the relating by the brain of patterns of input and output signals. Then, if language-learning establishes in me a relation between words (such as the words "telling the truth") and these patterns of output, I can say "I was telling the truth just now". The ability to say what one is doing is blind in a certain way, because the brain's discrimination is purely functional and is not dependent on the recognition of qualitative similarities and differences. Thus I can say what I know, and then go on to say that I have just said what I know, or in some other way I can act on the knowledge that I have just said what I know, without needing to have any criterion, as Wittgenstein would say, for this knowledge. I am enabled to report this knowledge or otherwise act on it, not by perceptual discrimination of qualitative conditions but by the programmed discriminations of neural functions.

Since the behaviour of saying what one knows has a physical characterization in terms of neurally stored information, and since my knowledge that I am behaving in this way depends on a neural function that discriminates this behaviour, it must be possible for a neural malfunction to occur analogous to the neural malfunction that produces pain "in" the amputated foot. Just as there may not be injury or even a bodily part to be injured when I know I have a pain (when I seem to have an injury), so there may not be actual stored information, information that has "passed the tests", when I say what I know - or seem to say what I know. This hallucination does in fact occur occasionally, particularly on waking from anaesthesia or a troubled sleep. One awakens with the "solution" to a nagging problem or the world's dilemmas only to realize a moment later that some nonsensical or fatuous message has arrived at the awareness line with false credentials.

This raises another point. When I say what I know (even in an hallucination of the type just described), surely I am aware that what I am saying is knowledge right when I say it, or even before I actually say it; it is not simply that afterwards I can say I was saying what I know. I am not denying that one can hesitate before one speaks, or that one can say something first to oneself and then aloud. If I know that Smith is here, I can say "Smith is here" or "I know this: Smith is here", or "Smith is here.

I know that for a fact". As soon as I can say "Smith is here" I can say I know it; if I postpone the utterance "Smith is here" for a second, I can preface it with "This I know: ... " or some similar fanfare.

CHAPTER 13.

Conclusions: Science and Language.

81. Loose Ends.

The paramount mystery of the mind is awareness or consciousness. The explanation of this phenomenon has required the analysis of many other matters. An analysis of perception was needed to lay the ghostly images of awareness. An analysis of reasoning was required to banish the persistent notion of awareness as an arena for mental action. The relationship between awareness and intention had to be examined to rid awareness of still another connotation: the wellspring of volitions. The concepts of content and infallibility, which are so bound up with awareness, demanded the chapters on content and knowing.

Taken together, these analyses amount to an almost complete physicalism. There are some loose ends, however, which must be acknowledged. The phenomena discussed so far have been the phenomena that have received the most attention from philosophers of mind, and have produced the most puzzles. But are they the major phenomena of mind?

There still remain certain human characteristics and capacities which are thought by many to be quintessentially spiritual and obviously beyond the grasp of the arm of physical science: humour, appreciation of beauty, romantic and spiritual love, and guilt, to name a few. That overworked fellow, the man in the street, might very well admit that animals and machines could be aware, could reason, intend and know, but he would insist that an animal or a machine could not fall in love, or thrill at the beauty of spring, or possess creative genius.

Just why the philosopher of mind should fix his attention almost exclusively on the less exalted mental phenomena is in itself an interesting question. One might think, for example, that whenever a dualist was faced with a barrage of troublesome criticism, he would - instead of qualifying and expanding his doctrines to meet the criticism in the picayune manner of philosophers - simply run out his big guns: if I have no soul or ego or mind, how can it possibly be true (as it so obviously is) that I can commune with beauty, be desolate with guilt, and love my neighbour? But philosophers do not do this; the rules of the philosophy game seem to be that disputes over problems of mind are to be settled by reference to the more workaday phenomena or not at all.

The obvious reason for the restrictions of the philosophy game is that the concepts involved in the spiritual life are just too woolly to make sense of. Philosophers have learned from experience that attempts to pin down the grand concepts of the spirit inevitably end in failure, so in the true spirit of science they turn first to what is relatively clear - or at least superficially unambiguous.

There are other considerations that hold more specifically for this thesis. There is a widely shared view that to the extent that love, creativity, humour, and so forth are real phenomena, they are somehow special constructions, combinations, or characteristic figurations of the less exalted phenomena - and nothing more. This view has a philistine ring to it, but it is certainly defensible, and no doubt goes a long way toward accounting for philosophers' impatience with those who bring love, God, and beauty into discussions of mind. This is the view the physicalist must take, but the actual reduction of descriptions of the spiritual life to the physical vocabulary of this thesis, while perfectly possible, would be tedious, unilluminating, and quite unpleasant. The reader who is sympathetic to my treatment can quite easily fill in the details if he is interested; there are no surprises and no serious obstacles. The reader who is antagonistic is unlikely to find any insuperable obstacles for me in this area - if he has not already found some in the preceding analyses. Producing the reductions would be unpleasant just because it would seem too philistine. The hypothetical physiological story behind the creation of Hamlet, for example, would hardly advance our understanding of great art, and would otherwise simply assault the sensibilities. One wants to object: "Is nothing sacred?", and the physicalistic reply is that nothing mental is intrinsically sacred, but since nothing of particular philosophic or scientific importance hinges on the spiritual life, it can be left inviolate.

Not strictly part of the spiritual life and also a subject of some philosophic interest is emotion, but here again the explanation of emotion in the terminology developed would be fairly dull and repetitive. Furthermore, psychoanalytic theories of emotion abound, and most of these would be physicalistic - if the authors gave the matter any thought. From the philosophic standpoint, Ryle's geography of emotions in the Concept of Mind seems to me both fundamentally sound and philosophically neutral. Dividing the field into inclinations, moods, agitations, and feelings, Ryle goes on to describe how love can be an inclination or an agitation, how grief is affection blocked by death, how moods differ from inclinations in monopolizing their effect on behaviour, how some agitations are frustrations of inclinations, and so forth. All of this he ties to his distinction between dispositions and occurrences, and it is here that physicalism comes in. If emotions are no more than dispositions and occurrences, there should be no theoretical snags in giving them physical embodiments on a par with such dispositions as brittleness and such occurrences as static in a radio. In fact, the search for generalized physical conditions in the brain as bases for behavioural dispositions is proving to be a fruitful, if not the most fruitful, area of neurological research.

But before accepting Ryle's analysis just as it stands, I would like to defend him on one easily misunderstood point. Ryle is often criticized for holding the view that dispositions need not be explained by reference to some existing condition. Ryle is held to be saying that true <u>hypothetical</u> statements about dispositions - like "If John Doe is addressed in French, he will respond appropriately in French", a hypothetical about the disposition of knowing French - do not point to the truth of some categorical statement about some existing state of affairs (in the brain or the mind, for example). When one knows French, <u>all</u> that is true is that a number of hypotheticals are true; no condition or state of affairs is the condition of knowing French.

Now if Ryle does hold this view, then he is not only wrong, but quite gratuitously wrong; the view is not necessary to the rest of his position. Geach points out the absurdity in the imputed view:

> A physicist would be merely impatient if somebody said to him: "Why look for, or postulate, any actual difference between a magnetized and an unmagnetized bit of iron? Why not just say that if certain things are done to a bit of iron certain hypotheticals become true of it?" He would be still more impatient at being told that his enquiries were vitiated by the logical mistake of treating "x is magnetized" as categorical, whereas it is really hypothetical or semi-hypothetical.[1]

It sometimes seems that Ryle may hold this view, but it is never explicit. He makes the lesser, and perfectly sound, points that a hypothetical is not a categorical, and that one's knowledge of the truth of a hypothetical does not depend on knowledge of the truth of any categorical <u>covering</u> statement. Whether or not there <u>must</u> be actual persisting differences behind all regularities of occurrences is a difficult metaphysical question not directly considered by Ryle (or by Geach). There is certainly every reason to believe that in the case of regularities in human behaviour (as in regularities in the "behaviour" of magnetized metals) there are governing structures to be found. The point to

1 <u>Mental Acts</u>, p. 6.

be made - whether Ryle makes it or not - is that right now all we know about most behavioural dispositions are the hypotheticals that are true; it is not the case that we know the nature of some (psychic or mental) condition that _is_ vanity, anger, intelligence, or knowing French, and hence when science does come up with some physiological conditions governing these dispositions, there will not be any question of whether the physical condition is _identical_ to some previously known or observed condition called anger or knowing French. Ryle is perfectly right in insisting that knowing that some people are vain is not knowing that some particular condition or quality is shared by these people. If he is saying that there is no shared condition ever to be discovered, then he is making a statement that may some day be refuted - but he does not seem to be saying this. And besides, the statement may well be true. Human brains are so different in minute structure that it may prove that the neural characteristics that determine vain behaviour in one person are completely unlike the neural characteristics in each other vain person, so that unlike the case of magnetized metals, there is nothing general that can be said about the internal condition of vain people, except that the conditions govern similar behaviour. Vain people may well be alike _only_ in their vain behaviour, and not in sharing any internal quality or condition of vanity. The regularities that science would then seek to explain would be regularities only in each individual, and even in the individual the conditions might change over a period of time. Just as an apple and a pillar box may both be red without sharing any determining reflective property, two vain people may behave vainly without sharing any determining physical characteristics. (See §27.)

If emotions are seen as dispositions, there is no problem about how one can be aware that he is angry or depressed. One discovers oneself acting angry and depressed just as one discovers that another is acting angry and depressed: by observation. The angry man has one evidential advantage over his observer, however, in that he may become aware of angry thoughts of which he inhibits the expression - either in words or in action. The "anger" remains when the external signs are suppressed, but one does not intuit some special presence in the psyche that is anger; one simply notices that one is going to some pains to inhibit angry actions and angry speech.

There are other events, qualities, and actions that are ordinarily seen as mental, but they are mental in about the same way playing chess is mental; mental activities are involved in them, but they are not separate mental phenomena. Doubting is one example; deceiving oneself or others is another. Insofar as there are philosophical problems with these, they are not problems in the philosophy of mind. Without "minds" things could not be ambiguous or interesting, but ambiguity and interest are not mental phenomena to be explained by the physicalist, or by any other philosopher of mind.

52. <u>Is This Science Fiction?</u>

A large part of the exposition and argument in this thesis has been about neurology, psychology, and cybernetics. Not only have scientific views been described, they have been edited, criticized, rejected, and elaborated on. The concept of the awareness line, the arguments on content and function, and on the non-imagistic nature of perception - all at least <u>quasi</u>- or <u>meta</u>-neurological ideas - are my own inventions. And I am not a scientist. Who am I, then, to write on such matters, and do my arguments have any claim to be heard, let alone accepted?

The systematic description and explanation of mental phenomena presented here might be defended as part of a long philosophical tradition. Lucretius built his minds of slippery, fast-flowing atoms, Descartes located interaction in the pineal gland, Hume reduced the mind to impressions and ideas. More recently, the systematic nomenclature has turned to sense-data, sensa, sensibilia, acts of judgment, mental causes, raw feels, noemata, Gestalts, ids and super-egos. Are my message-carrying signals, storage patterns, and awareness lines any more irresponsible concepts than these? But I have held that these recent concepts are by and large philosophically irresponsible, so I cannot defend myself by appealing to tradition or recent practice.

My defence is that somebody has to do the job I have attempted here. Two sorts of mistakes are persisting in the efforts to understand the mind: neurologists and psychologists are making philosophical mistakes, and philosophers are making both philosophical and scientific mistakes. The professional neurologists who are amateur philosophers have tried their hand (W. Russell Brain, J.Z. Young, and Ragnar Granit, to name the best), but their philosophy has not been up to the standards of their science. Perhaps the bridge must be built from the other side.

My method has been to adjudicate between various philosophical and physiological hypotheses on the basis of considerations from three quarters: philosophy, psychology, and neurology (including cybernetics). This is surely better than the standard philosophical practice of simply ruling out physiological solutions on philosophical considerations alone. Many philosophers believe that there is a large realm of questions in the philosophy of mind to which physiology and psychology are totally irrelevant. I hope that although all the answers provided here may not be right, I have shown that these answers are neither irrelevant nor absurd.

From the scientific point of view, my systematic explanations can be criticized on two fronts. First, they are underdetermined by experimental data. Second, they are relatively vague and incomplete. The answer to the first charge is that my hypotheses are presented as speculation. Not as wild speculation, but as "good chances" or "best bets" or even, in some cases, as predictions of what will in time be satisfactorily demonstrated by science. Admittedly, this disavowal of having proved my scientific case places what I have said somewhat out of reach of immediate disconfirmation on scientific grounds. The philosophic case is more concerned with possibilities than actualities (at the moment), and other sorts of criticisms are available from that quarter.

The answer to the second charge is related to the first. Until such time as the scientific results come in, it is better for those who are considering possibilities to avoid taking stands where a variety of different details could fill out the description. Neurones and the other equipment of the brain have been considered more or less as "black boxes". The fine mechanical details of their operation have been left unconsidered. If this is not a complete story, it is an advance on previous stories. While many philosophers treat either the human body or the human mind as a black box with certain outward behaviour, my account carries the examination into the interior. At least the unexplained operations of my black boxes, neurones, are not of the order of mystification of such unexplained operations as intending, recognizing, initiating, deciding, imagining and dreaming. If questions like "How can we infallibly peruse our sense-data?" can be reduced to questions like "How can a neurone change its threshold?" and "How are feedback mechanisms in the brain organized?", a very definite gain has been made.

A further defence of the general story at the expense of the fine details is that we now know that the brain is so complicated that a rigorous step-by-step production of the general story from the details will not be forthcoming in the indefinite future. In the meantime, under-determined general stories are useful. In the course of explanation I have discussed two activities which will probably never occur: the denial of the subject veto by an immense computer-aided operation, and the ascription of content to high level neural signals. The fact that these activities are in principle possible is of as much importance to the philosophical problems of mind as the fact of their actual occurrence would be. As Arbib says, "We may not yet have modeled the mechanisms that the brain employs, but we have at least modeled possible ones, and that in itself is a great step forward."[1]

At this point it may be well to consider the problems of parapsychology, since it might be held that the findings on ESP and psychokinesis simply rule out the sort of physicalistic explanations presented here. J.B. Rhine says, at one point, that for parapsychological phenomena, "physical explanation seems to be clearly excluded".[2] But Rhine does not mean by this that the explanation must be spiritual; he means that the part of the physical world usually reserved for explanations of human activity, in the main physiology and biochemistry, will not do for ESP. He fully recognizes, as others do not, that even if some new force must be postulated to account for these phenomena, the force will be a physical force, though perhaps quite unlike

---

1 Op. cit., p. viii.
2 "Parapsychology and the Nature of Man", in Hook, op. cit., p. 74.

other physical forces.

However, the findings of Rhine and others, if they are finally accepted, do point to a difficulty for the neural net type of theory. No allowance has been made for neurones to be any kind of interpersonal transmitter, and furthermore, the only sense that has been made of neuronal signals having content is tied strictly to intra-cerebral functions. If neural signals have content only insofar as they function to determine behaviour in one individual, it is hard to see how content could be transmitted from one individual to another. Aside from parapsychological findings, there is every reason to believe that brains do not in any way "speak the same language".

Psychokinesis and clairvoyance present further difficulties, and if and when the existence of these phenomena is firmly established, a physicalistic theory of mind must take them into consideration. But the champions of rival theories of mind should find no comfort in these possibilities. If there is psychokinesis and ESP, their existence presents just as many problems to the epiphenomenalist or the interactionist. Saying that minds can communicate within the mental sphere alone without need of physical interaction does not explain a thing. It is merely rephrasing in suspect terminology the sort of hypotheses that are sometimes entertained by the researchers in this field. A _deus ex machina_ is not an explanation but merely a roundabout admission that one has no explanation.

53. _Arguments About Language._

The treatment of language throughout the thesis has been a variation on the "ordinary language analysis" method. By analyzing the ordinary verbal behaviour surrounding a phenomenon before attempting to describe it, one ensures that one is asking the right questions, and just as important, that those who ask the wrong questions are led to accept the right questions. By treating complete utterances in their contexts as the objects of scrutiny, rather than plunging into attempted explanations of just what things perceptions, sensations, intentions, volitions, and reasons _are_, one ensures that one's ontology is not inflated by linguistic oddities and metaphors. Behind successful or appropriate talk there must be actual differences in states of affairs, but there need be differences only behind the differences in whole utterances, and not necessarily behind the grammatical distinctions within the utterances. Our ordinary systems of grammar are not cages within which reality is preserved.[1]

1 Cf. Quine, _Word and Object_, Chs. 2 and 3.

My method differs from the standard one in treating such linguistic analysis as a preparation for doing philosophy, rather than as the end of philosophy. A simple point that often seems lost on the analyzers of ordinary concepts is that the phenomenon of perception and the concept of perception are two different things. After one has decided what the ordinary concept of perception (or intention, or mind, or dreaming) is, there is still room to ask whether this is a good concept. The concept that one can piece together from the vagaries of ordinary usage may be inconsistent, ontologically suspect, or an amalgam in one notion of several separable notions - as was the case with the concept of awareness, for example. So my method has been first to conduct a relatively cursory examination of ordinary usage, and then separate what is separable, group what goes together, and discard what is plainly irrelevant - all with an eye on the phenomenon itself. This leads to the production of new, non-ordinary uses for ordinary words, and gradually the questioner is drawn away from the folk philosophy of his inherited language to a set of more manageable concepts.

Is this not just what the questioner wants? When a person wants to know what seeing is all about, is he satisfied to learn how he in fact uses the word "see"? The knowledge of this ensures that his usage will be in static conformity with his peers, but presumably he asked the question because he felt a certain confusion in the word, not because he felt uncertain when it would be proper to use it. The method of linguistic analysis taken neat serves only to describe the confusion clearly, but describing confusion is not clearing it up.

Jerome Shaffer has suggested that ordinary language is fundamentally dualistic with regard to the mind-body problem.[1] But even if one accepts this result of ordinary language analysis, one cannot infer that dualism is right. All one can infer - at most - is that the unreflective man in the street presumes that dualism is right. Smart suggests that if Shaffer is right, then revision of ordinary language might be appropriate.[2] Revision is certainly possible, but except for theoretical purposes, clarity on these matters is not very important. Provided a bridge exists between the ordinary and non-ordinary use, there is no need for full-scale reform.[3]

Philosophy can be seen as a democratic institution; it accepts the burden of answering the questions of the man

---

1 "Could Mental States be Brain Processes?", Journal of Philosophy, 1961, pp. 813-22.
2 Op. cit., p. 98.
3 Cf. Quine, op. cit., §45, "The Double Standard".

in the street to <u>his</u> satisfaction. Meanwhile the scientists may answer the special questions of their elite, questions about phenomena occurring in the world. Wherever the explanation of occurring phenomena comes into the philosophical answers to questions, the philosopher's task is either to bring the man in the street to science, or to bring science to the man in the street.

The values and shortcomings of linguistic analysis as an end in philosophy are nowhere more evident than in Anscombe's <u>Intention</u>. By examining the rôle of whole utterances in our talk about intentional actions, she manages to sweep away the supposition that when one says what his intentions are, there must be an intention (a thing or event) that one is describing. She describes not intentions, but the situations in which we speak of intentions, and this leaves the way open to answer the question: What is the difference between an action we call intentional, and an action we call unintentional? The answer to this is not just that we call actions intentional when we also say such and such; there is, and must be, another sort of answer. She never asks if the way we use "intentional" serves consistently to pick out something of importance. She might well have taken her cue from Ryle and called her book <u>The Concept of Intention</u>, for although she discusses the ordinary concept, she never gets around to discussing the phenomenon.

It might be thought that that is just the dividing line between philosophy and science: philosophy treats of concepts, and science treats of phenomena. But I have attempted to show that one cannot examine concepts critically without reference to phenomena, and if one cannot examine concepts critically, one may well without knowing it be examining nonsense. One examines the ordinary concept of intention because it is important in human affairs; a great deal quite literally hangs on it. Having decided just what the ordinary concept of intention is, one cannot then justify its use by pointing out that it is important in human affairs just because we use it, because it is ordinary. One criticizes idealists and Continental philosophers and the like for systematizing exotic nonsense, but by not turning to the phenomena, the analyzers of ordinary language are in danger of systematizing ordinary nonsense, a small gain.

54. The Ghost in the Machine.

The little man in the brain, Ryle's ghost in the machine, has been rightly an object of ridicule for years. My intent in this thesis has been to show that his removal from the brain is a radical and far-reaching theoretical step. When one gets rid of the little man, in all his disguises as Ego, agent, sense-datum peruser, source of volitions, Smart's and Feigl's stimulus-checking mechanism, and introspector, one must also abolish the objects with which he was supposed to work. With no little man, there is no room for mental images, hence no room for sense-data, noemata, subjective or phenomenal space, or even raw feels. With no such objects of perception or sensation, there is no room for special emergent qualities of these objects. With no emergent qualities, there are no problems about their ultimate homogeneity, and hence no problems about ultimate divisions between the phenomena of the universe. Gone too are volitions, and their problems. Gone is the problem of infallible intuition of experience. Gone are mental acts. Not only must all these exotic entities for theory be banished, but all their everyday counterparts as well; thoughts, ideas, pictures in the imagination and acts of will. These may remain for use in ordinary affairs, but not for use in theorizing. Bring any of these objects back into the fold of the philosophy of mind, and the little man must be brought back with them. Raw feels must be felt, sense-data must be had and intuited, acts of will must be performed, and the only candidate for these offices is the little man in the brain. One is a physicalist or one is an animist; there is no middle ground. Epiphenomenalism makes no sense unless an epiphenomenal introspector is posited along with the parade of mental events. Qualities, whether describable or ineffably "raw", have no place in awareness unless one is also willing to posit a thing to sense these strange qualities.

The alternative to the ghost in the machine is a view of man as the possessor of an immense system of information processing and controls. For ease of understanding, one interprets the activities of these systems as endowed with content. This interpretational world then encompasses the problems of teleology and intentionality; it is a world of purpose and meaning. We now know that the human brain is an organ of vast powers as a receiver of information and a director of behaviour. Is the brain not a grand enough thing to replace the mind?

BIBLIOGRAPHY

ALLPORT, Gordon — "Eidetic Imagery", British Journal of Psychology, Vol. 15, 1924

ANDERSON, Alan Ross — Minds and Machines, New Jersey, Prentice Hall, 1964.

ANSCOMBE, G.E.M. — Intention, Blackwell, Oxford, 2nd edition, 1963.

ARBIB, Michael A. — Brains, Machines and Mathematics, New York, McGraw-Hill, 1964.

AYER, A.J. — The Problem of Knowledge, Penguin Books, 1956.

- - — The Concept of a Person, London, Macmillan, 1963.

BAIER, Kurt — "Smart on Sensations", Australasian Journal of Philosophy, 1962.

BENNETT, Jonathan — Rationality, London, Routledge & Kegan Paul, 1964.

BRAIN, W. Russell — Mind, Perception and Science, Blackwells, Oxford, 1951.

- - - — "Some Reflections on Brain and Mind", Brain, 1963.

- - - — "Space and Sense Data", British Journal of the Philosophy of Science, 1960.

BRANDT, Richard B. — "Doubts about the Identity Theory" in Hook, Dimensions of Mind.

BURT, Cyril — "The Concept of Consciousness", British Journal of Psychology, 1962.

CHISHOLM, Roderick — Perceiving - A Philosophical Study, Cornell and Oxford, 1957.

DANTO, Arthur C. — "Consciousness and Machines" in Hook, Dimensions of Mind.

ECCLES, Sir John — "The Synapse", Scientific American, Jan. 1965.

EDWARDS, Elwin — Information Transmission, London, Chapman & Hall, 1964.

FEIGL, Herbert — "The 'Mental' and the 'Physical'" in Minnesota Studies in the Philosophy of Science, Vol. II, 1958, Minneapolis.

FIRTH, Roderick — "Sense-Data and the Percept Theory", Mind, 1949.

FREGE, Gottlob — The Foundations of Arithmetic, 2nd revised edition, translated by J.L. Austin, New York, Harper Torchbooks, 1953.

GEACH, Peter — Mental Acts, London, Routledge & Kegan Paul, 1957.

GEORGE, F.H. — The Brain as a Computer, Oxford, 1961.

\- - Cognition, London, Methuen, 1962.

GRANIT, Ragnar A. — Receptors and Sensory Perception, Yale, New Haven and London, 1956.

GRICE, A.P. and STRAWSON, P.F. — "In Defense of a Dogma", Philosophical Review, 1956.

GUSTAFSON, Don F. — "On the Identity of Theory", Analysis, 1963.

HESS, E.H. — "Space Perception in the Chick", Scientific American, 1956.

HOOK, Sidney (ed.) — Dimensions of Mind, N.Y.U., New York, 1960.

HUBEL, D.H.	"The Visual Cortex of the Brain", Scientific American, Nov. 1963.
HUBEL, D.H. and WIESEL, T.N.	"Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex", Journal of Physiology, 1962.
HUSSERL, Edmund	Ideas, London, George Allen & Unwin Ltd., 1931, translation of "Ideen", 1913.
HUSSERL, Edmund	Logische Untersuchungen, Halle, 1922.
JEFFERSON, Geoffrey	"The Mind of Mechanical Man", British Medical Journal, June 25, 1949.
JOSKE, W.D.	"Sensations and Brain Processes: a Reply to Professor Smart", Australasian Journal of Philosophy, 1960.
KINSBOURNE, M. and WARRINGTON, Elizabebh K.	"A disorder of simultaneous form perception", in Brain, 1962.
KONORSKY, Jerzy	Conditioned Reflexes and Neuron Organization, translated by S. Garry, Cambridge, 1948.
LASHLEY, Karl	Brain Mechanisms and Intelligence, Chicago, University of Chicago Press, 1929.
- -	"The Problem of Serial Order in Behavior" in Lloyd A. Jeffress (ed.), Cerebral Mechanisms in Behavior, The Hixon Symposium, 1951, New York, Wiley.
LETTVIN, J.Y., MATURANA, H.R., McCULLOCH, W.S. and PITTS, W.H.	"What the Frog's Eye Tells the Frog's Brain", in Proceedings of the Institute of Radio Engineers, 1959.

LETTVIN, J.Y., MATURANA, H.R., McCULLOCH, W.S. and PITTS, W.H. — "Two Remarks on the Visual System of the Frog" in Rosenblith, W. (ed.), Sensory Communications, .. 1961.

LEWIS, C.I. — Mind and the World Order, New York, Scribners, 1929.

- - — An Analysis of Knowledge and Valuation, La Salle, Indiana, Open Court, 1946.

LURIA, A.R., PRAVDINA-VINARSKAYA, E.N. and YARBUS, A.E. — "Disorders of ocular movement in a case of simultanagnosia", Brain, 1963.

McCULLOCH, W.S. and PITTS, W.H. — "A Logical Calculus of the Ideas Immanent in Nervous Activity", Bulletin of Mathematical Biophysics, 1943.

MacKAY, D.M. — "Toward an Information-Flow Model of Human Behaviour", British Journal of Psychology, 1956.

MacNICHOL, Edward F. Jr. — "Three-Pigment Color Vision", Scientific American, Dec. 1964.

MACE, C.A. — "Introspection and Analysis" in Max Black (ed.), Philosophical Analysis, Ithaca, Cornell University Press, 1950.

MARGOLIS, Joseph — "Intention, Consciousness, and Action", Methodos, XIV, 1962.

MILLER, G.A., GALANTER, E. and PRIBRAM, K.H. — Plans and the Structure of Behavior, New York, Holt, 1960.

MUNTZ, W.R.A. — "Vision in Frogs", Scientific American, March, 1964.

NEISSER, Ulric — "Visual Search", Scientific American, June, 1964.

PITCHER, G. — "Sensations and Brain Processes: a Reply to Professor Smart", _Australasian Journal of Philosophy_, 1960.

PRICE, H.H. — _Perception_, London, Methuen, 1932.

PUTNAM, Hilary — "Minds and Machines" in Hook (ed.), _Dimensions of Mind_.

QUINE, Willard Van Orman — "Two Dogmas of Empiricism", _Philosophical Review_, 1951, reprinted in Quine, _From A Logical Point of View_, Harvard, 1953.

\- - - - — _Word and Object_, Cambridge, Mass., The Technology Press of M.I.T. and John Wiley & Sons, Inc., 1960.

RAPOPORT, Anatol — "Technological Models of the Nervous System", _Methodos_, 1955, reprinted in Sayre and Crosson, _The Modeling of Mind_.

RHINE, J.B. — "Parapsychology and the Nature of Man" in Hook (ed.), _Dimensions of Mind_.

RUCH, T.C. — "Motor Systems", in S.S. Stevens (ed.), _Handbook of Experimental Psychology_, New York, Wiley, 1951.

RUSSELL, Bertrand — _The Analysis of Matter_, London Routledge, 1927.

RUSSELL, W. Ritchie — _Brain - Memory - Learning_, Oxford, 1959.

RYLE, Gilbert — _The Concept of Mind_, London, Hutchinson, 1949.

SAYRE, Kenneth — "Human and Mechanical Recognition" in Sayre and Crosson, _The Modeling of Mind_.

SAYRE, Kenneth and CROSSON, Frederick J. (eds.) — _The Modeling of Mind_, Notre Dame, Indiana, 1963.

SCRIVEN, Michael — "The Mechanical Concept of Mind", Mind, 1953, reprinted in Anderson, Minds and Machines.

SELLARS, Wilfrid — Science, Perception and Reality, London, Routledge & Kegan Paul, 1963.

SHAFFER, Jerome — "Could Mental States be Brain Processes?" in Journal of Philosophy, 1961.

SHORTER, J.M. — "Imagination", Mind, 1952.

SIMMEL, Marianne L. — "Phantom experiences following amputation in childhood", Journal of Neurology, Neurosurgery and Psychiatry, 1962.

SKINNER, B.F. — Science and Human Behaviour, New York, Macmillan, 1953.

SMART, J.J.C. — "Sensations and Brain Processes", Philosophical Review, 1959.

- - — "Sensations and Brain Processes: a Rejoinder", Australasian Journal of Philosophy, 1960.

- - — "Further Remarks on Sensations and Brain Processes", Philosophical Review, 1961.

- - — "Brain Processes and Incorrigibility", Australasian Journal of Philosophy, 1962.

- - — Philosophy and Scientific Realism, London, Routledge & Kegan Paul, 1963.

SMYTHIES, J.R. — "Analysis of projection", British Journal of the Philosophy of Science, 1954.

- - — The Analysis of Perception, London, Routledge, 1956.

SMYTHIES, J.R. "The Problem of Perception", <u>British Journal of the Philosophy of Science</u>, 1960.

SPRAGUE, Elmer "The Mind-Brain Problem" in Hook (ed.), <u>Dimensions of Mind</u>.

STEVENSON, J.T. "Sensations and Brain Processes: A Reply to J.J.C. Smart", <u>Philosophical Review</u>, 1960.

TAYLOR, Charles "Phenomenology and Linguistic Analysis" in <u>Proceedings of the Aristotelian Society</u>, Supplementary Volume, 1959.

- - <u>The Explanation of Behaviour</u>, London, Routledge & Kegan Paul, 1964.

TURING, A.M. "Computing Machines and Intelligence", <u>Mind</u>, 1950, reprinted in Anderson, <u>Minds and Machines</u>.

VAN SENDEN, M. <u>Space and Sight</u>, London, Butler & Tanner, 1960, translation of <u>Raum- und Gestaltauffassung bei Operierten Blindgeborenen</u>, 1932.

WARNOCK, G.J. "Seeing", in <u>Proceedings of the Aristotelian Society</u>, 1954-55.

WEVER, E.G. "Figure and ground in the visual perception of form", <u>American Journal of Psychology</u>, 1927, Vol. 38, quoted in George, F.H., <u>Cognition</u>, 1962.

WIENER, Norbert <u>Cybernetics</u>, 2nd edition, 1961, New York, M.I.T.

WITTGENSTEIN, Ludwig <u>Philosophical Investigations</u>, translated by G.E.M. Anscombe, Oxford, Basil Blackwell, 1958.

YOUNG, J.Z. <u>A Model of the Brain</u>, Oxford, The Clarendon Press, 1964.