

# On Body and Mind

**Michael Polanyi**

[This article originally appeared in *The New Scholasticism* 43:2 (Spring 1969): 195-204. It is put on WWW with kind permission of the American Catholic Philosophical Association and John C. Polanyi. The 1969 publication indicates that the essay was adapted from a recording of an address to psychiatrists at Duke University on April 16, 1964.]

In November 1963 there has appeared in *Encounter* a paper by Rattray-Taylor of the University College, London, under the title *The Age of Androids*. It gave a lucid account of the way computers simulate the mental acts of human beings and raised the question, whether we will have to attribute to a computer the powers of reason and the feelings of consciousness which we attribute to human beings, provided the computer gives the same responses as a human being, to whom we attribute these states of mind. The question is an acute one and worth considering, because it is established by the theoreticians of computers, that whatever specifiable response you want to elicit from a computer to a specifiable input, you always can construct one to achieve that response. The question raised by Rattray-Taylor, and raised by many others in similar ways, amounts then to this: Is it possible to give a specific definition of a mental performance, like thinking or understanding, or of human feelings, in terms of a performance by which we judge the presence of these states of mind.

This question has hardly ever been discussed in detail, but some sixteen years ago the great mathematician von Neumann, speaking at the Hixon Symposium of 1948, rejected the assumption that there might be performances which we cannot define in a finite number of words, as a mysticism which we must not seriously countenance. This view seems generally shared today.

I shall suggest, on the contrary, that all communication relies, to a noticeable extent on evoking knowledge that we cannot tell, and that all our knowledge of mental processes, like feelings or conscious intellectual activities, is based on a knowledge which we cannot tell. And if we do recognize mental processes by noticing things we cannot tell, it follows that it is not possible to construct a machine which would give the same responses as those by which we recognize these mental processes.

The facts I am referring to are commonly known. Edna Heidbreder (1947) has concluded from her experiments that "we very often fail to define a concept adequately, though we are entirely capable of using it." Everybody can confirm this from his own experience.

Common experience also tells us that in teaching we rely on an intellectual effort of the learner for recognizing that which we are conveying to him. This is particularly true when we teach medical diagnostics or the recognition of botanical, zoological, histological or geological samples in practical classes. It is also true in teaching to identify sensory qualities; for example, judging the character of a pulse, its fullness and various other qualities which we ascribe to it; and perhaps even more delicately, when teaching the reverberations on which we rely in judging a percussion that we are performing. All skilful performances can be taught only by relying on the pupil's effort to imitate and catch the knack of it, by the intellectual effort to find out how it is done. And in the end the pupil, in his turn, cannot tell how he performs the skill.

There is yet another connection in which we have to rely on an effort of the imagination in teaching and

learning. It is exemplified by the study of topographic anatomy. You can see various stages of dissection, in fact or in pictures, but you have to reconstruct by the effort of the imagination how the organs, how these several elements, hang together, so that you know your way about them. There are a number of branches of knowledge which are based on a similar understanding: architecture, machine structure, the structure of crystals, the stratification of geological sites.

We have here, then, a number of examples in which our teaching relies on the capacity of the learner to discover for himself a considerable part of that which we are trying to impart to him, and to this extent we are imparting to him something that we cannot tell, and which he, in his turn, then knows and cannot tell.

I pass on now to a less obvious field, but one well-known, namely, that of visual perception. It is based of course on producing a retinal image of an object on which we are focusing, but it is well known that the way this object will appear to us, will depend largely on marginal clues, which we may not notice at all. For example, it may depend on the variation of the amount of light which enters the eye as we focus the lens on variable distances, or on some slight detail at the edge of the visual field. A shadow seen from the corner of our eye, which shows that the sun is shining, will make us see an object at the center in a different way. These marginal clues are to a considerable extent unknown to us. They are unknown in the sense that we cannot tell that we are using them. We get to know them only by studying the effect of eliminating them.

There are also internal clues of perception, which determine the way we see things. Oculomotoric proprioception is of course essential and messages from the labyrinth also contribute effectively. And I would venture to go deeper than these internal clues and to bring in also the ensuing cortical processes as events that determine what we see and which we cannot tell.

There is an ambiguity in all these examples, between the two kinds of knowing which I have covered by the term "knowing." The object of our attention is always there and can be identified. This we know and can tell. But we cannot be sure in identifying the clues which shape the knowledge that is the subject of our attention, or the muscular elements which contribute to a skillful performance. These are the marginal elements, which *we know and may not be able to tell.*

I would say that these elements are *subsidiary* to what we are focally attending to, and that there is a *functional relation* (I am borrowing this term from Gestalt Psychology) between the marginal elements and the object to which we are focally attending. We are *subsidiarily aware of the marginal elements, with a bearing on that which is focally known.* This relation is the product of an integration which links the subsidiary elements to the focal center. Both the knowledge of these subsidiaries and of the integration which links them to the center is, or may be, *tacit*, that is, *such that we cannot tell it, except quite vaguely.*

Let me repeat once -more the nature of this functional re-lation. It consists in being aware of something for the purpose of attending to something else. We then may be aware of it only in terms of that to which we are attending from it. We may say that we attend thus from a clue to that to which it points. And we may note also that a clue may lead to a discovery and become part of that which will be discovered. We can also say that the clues have a *semantic function*, of pointing at something of which they are parts or prospective parts. They mean those things. What we see at the edge of our visual field and the proprioceptive oculomotoric stimuli which reach us, mean something in terms of visual perception.

This relationship of functioning and meaning causes a change in the appearance of the elements on which it operates. Take for example, the disparity of two stereo photographs, which we see--by using one eye for each of them--as a single spatial object. This is a transformation in the quality of subsidiary elements, when transposed into our focal seeing of something to which they contribute.

Finally, I would say that this tacit integration. is a *comprehension*. It may *produce* a comprehensive entity in the form of a skill or *recognize* the presence of a comprehensive entity, as we do when we recognize an object by visual; clues, both external and internal to us.

This is also how we recognize a physiognomy. We attend, from a large number of clues most of which we cannot identify, to the physiognomy which they indicate. All comprehensive wholes are known in this way. We attend *from* their parts *to* the whole.

The first point I want to derive from this analysis, is that we should look upon the body-mind relation in this context. We have seen that a perception is achieved, by attending to an object from the marginal clues and from internal clues, of which some are subliminal. I suggest now that we extend this scheme to include a subsidiary awareness of cortical processes, by recognizing that we are aware of these too in terms of that which we perceive. This offers a unified view of consciousness. It affirms that consciousness invariably attends *from* something subsidiarily *to* something focally, and that this includes the body-mind relation in its entirety.

We can extend this structure now to the comprehension of things outside. I have said that comprehensive entities are recognized by attending *from* their parts *to* the whole: This is then to use the parts in the way we use our body for attending to objects outside. In other words, it is to recognize a whole by interiorising its parts so as to attend from them to their joint meaning. Such an interiorization of parts I call indwelling.

Indwelling uses the parts of a thing we comprehend in the way we use our bodies in the process of perception. And consequently, a great deal may be subsidiarily known in identifying a comprehensive object that cannot be specified and may indeed be subliminal. This explains why there is so much that we know and cannot tell when identifying a person, or sensing his feelings and thoughts by his behavior.

Konrad Lorenz (1959) has made a contribution to this argument[[1](#)] by pointing out that a much larger number of items can be taken into account and far more rapidly included in a Gestalt integration of the kind I have been talking about, than by focally attending to several items. We can see this fact illustrated, if we think of the way we learn a skill by following the instructions of a manual. Suppose we learn to drive a motor car from the words of a manual, we shall have to practice its directions until we have completely interiorized them, and made them recede to the back of our minds. By such an interiorization we become capable of attending to a much larger number of items, and far more quickly, than we could while we were using the directions of the manual explicitly.

The opposite effect is also important. We can observe the effect of *exteriorization* when we attend to words or the elements of a skill which we had previously used subsidiarily. If we repeatedly pronounce a word, while watching our lips and tongue and the sound that we are making, the word will soon sound hollow and presently lose its meaning. And it is well known also that when we watch our fingers in playing the piano, this can paralyze us. The elements of a skill, which could be rapidly and successfully integrated, so long as they were interior to us, cannot be anything like as rapidly and successfully integrated once we are attending to them focally. They become meaningless, and paralyze us.

Before coming to the main question, I have yet to make a slight digression to tell you what a theory of knowledge in general looks like on this basis. We must recognize first of all that in spite of all my praises of tacit knowing, it is explicit knowledge which, by its immense powers, distinguishes us from the animals and grants us all our superiority in our intelligence over the animals. It is also true that explicit definitions explicit inferences operate with a minimum of indwelling. Yet the fact remains that all explicit argument must start from experience which is achieved by indwelling. All the input and output has the structure of indwelling of the very kind we use in perception. In the exact sciences these tacit roots the process of inference and the tacit

bearings of explicit theories, are comparatively trivial and therefore the exact mathematical sciences appear almost completely composed of theories which are impersonal. But remember what was said about the teaching of biological, psychological and similar subjects. Remember the intelligence and skill required in learning physiognomies and manipulations. Here we have comprehensions of great depth and amplitude which alone can provide us with an expert understanding of the experience to which our theories refer. In this case the explicit argument is merely a link between an input and an output which is based on the processes of tacit knowing that I have described. The explicit argument is external to this input and output.

This brings me to my first conclusion, which will dispose of the question with which I have opened this paper. No specifiable response can define the presence of sentient and reasonable human beings. We know a million times more about the appearance of sentience and reason than we can tell, let alone express in exact terms. Hence we cannot attribute sentience or reason to machines.

From the critique of machines we can pass on to a critique of neurology. Does our nervous system learn, recognize, patterns, think, feel? Can it be conscious? The conscious operation of the nervous system were established by the neurologists by accrediting sensory, emotional or other experiential reports. These reports had no other meaning than that which they possessed by their hearing on the presence of sensory, emotional and other mental qualities ascribed to the subject. Both input and output of a neural process are meaningless, except for having a bearing on the mental experiences of the subject. There are neural processes which have no such bearing, because they are purely physiological functions--for example the functions of the sympathetic system. But I am concerned with the analysis of conscious processes. I can repeat therefore that the bearing by which we understand both the input and the output of a neurological process must be established by ourselves, by our interpretation of the behavioral and neural signs of this input and output. *The neural functions supply these signs, but they do not supply their interpretation.* Since this interpretation forms no part of the nervous system, the system cannot be said to feel, learn, reason, et cetera. These are experiences or actions of the subject using his own neural processes.

But how can it be that the nervous system has this duality, that it *accounts* for mental experiences, but it does *not actually have these experiences*. The simplest way of seeing this situation is to think of the optical functions of the nervous system, including its central parts. We cannot see what the subject sees by the nervous processes of optical perception taking place in his neural system. The powers exercised by the subject when using his nervous system are superior to ours when observing his nervous system, because he is dwelling in the system in a way which is not accessible to us as observers of the system. This inferiority of the observer's powers is of the same kind as that of the speaker who is paralyzed by attending to his lips and tongue and the sound of his voice. We can share the situation of a person sufficiently to observe his mental processes but we cannot share his situation as he does dwelling in his own nervous system. Hence, this system does not afford any conscious experience to us as observers, as it does to the subject who fully dwells in it.

The logic of tacit knowing offers in principle a standpoint for evaluating the program of behaviorism. This program sets out to replace all reference to mental states by descriptions of the behavior by which these states are known to us. There could be no objection to the behaviorist program, if the piece of behavior which correspond to the presence of a mental state could be focally known. But they are in fact known subsidiarily, as *clues to mental states*. We could never get to know them at all, except by attending to the mental states on which they bear. We are attending *from* this external behavior *to* the mental states on which it bears and could never get to know these external signs in any other way.

Let us recall what was said about the semantic aspect relationship, namely that the whole to which we "are attending from its particulars, is their meaning. We have integrated then the particles with this meaning in mind and having done so, we can try to attend to them separately in themselves. But such specification, if we could

carry it out, would destroy their meaning. We would have meaningless elements, meaningless pieces which we could no longer identify at all as forming part of a behavior representing a state of mind.

We can in fact continue to identify them only because we cannot shift our attention wholly to these elements. Many of the particulars will not be specifiable at all and some will be subliminal. Moreover, the principles of integration will be those of a physiognomy or a skill, of things that can be described only to people who are already familiar with them. These subliminal traces and undefinable principles we will not be able to identify and they will be left lurking at the back of our mind. So the denaturing of the pieces of behavior, reduction to meaningless fragments, will not be complete.

We should note also that the sheer number of particulars and the speed of the rapidly changing integration of the external signs of a mental process, will make their focal observation in their ever changing configuration impossible. We will not really achieve their externalization, but will achieve something which is more difficult to define: namely, some semblance of the original program.

Just how much of the original program is thus carried out, I cannot analyze here much further. I will close therefore by repeating something I have written on this subject before.[2]

The present psychology of learning which strives for objectivity, achieves a semblance of its aim as follows. First, it curtails its subject matter to the crudest forms of learning. Second, it exploits the ambiguity of its supposedly impersonal terms, so that they will apply to the actual performance which is covertly kept in mind. I have called this a pseudo-substitution. It consists in using objectivist terms which are strictly speaking nonsensical, as pseudonyms for the mentalistic terms which they are supposed to eliminate.

There is an excellent paper by Chomsky, linguist and philosopher at MIT. It is an essay[3] on verbal behavior, in relation to a book called *Verbal Behavior* by B. F. Skinner. It contains a number of illustrations in favor of what I have just said, namely that Skinner uses terms like stimulus, control, response, and so on, in such a way that by their ambiguity they cover the mentalistic terms which they are supposed to replace. Chomsky shows that either you use these objectivist terms literally, then what is said is obviously false and absurd; or you use them as substitutes for the terms which they are supposed to eliminate, as pseudonyms of these, then you don't say anything else than what you would have said in mentalistic terms and the whole performance is empty.

This is too severe a judgment if taken to apply generally and I don't intend it to be taken thus. I want only to show the key to the critique of all attempts at defining mental processes in terms of the behavior by which we recognize them. Such definition cannot be carried out and what it actually says owes its meaning to an ambiguity by which it includes what it professes to replace.

### Notes

[1] Konrad Lorenz *Gestalt Perception as Fundamental to Scientific Knowledge, in General Systems*, ed., L. von Bertalanffy and A. Rapoport (Ann Arbor, 1962), p. 50.

[2] See, *Personal Knowledge* (London and Chicago, 1958), p. 369.

[3] This essay has since been published separately as a pamphlet.