The Structure of Consciousness
Michael Polanyi

[This essay originally appeared in *Brain*, Vol. LXXXVIII, (1965), pp. 799-810 (for the *Brain* home page, see http://brain.oupjournals.org/). It is put on WWW with kind permission from *Brain*, Oxford University Press and John C. Polanyi. The essay also was included in *Knowing and Being*, ed. M. Grene, London: Routledge, 1969, pp. 211-224. The text below follows the *Knowing and Being* publication which has a modified first paragraph.]

Sir Francis Walshe, in whose honour this essay was first published [1], has often spoken of the inadequacy of anatomic structures to account for the full range of mental actions; he insisted on the presence of integrative mental powers not explicable in these terms. Toward the end of this paper I shall give reasons supporting this view.

**Two Kinds of Awareness**

I shall start with an analysis of perception and shall arrive by successive generalizations of the result to a stratified structure of living things, which will include the structure of consciousness in higher animals.

Take a pair of stereoscopic photographs, viewed in the proper way, one eye looking at one, the other eye at the other. The objects appear then distributed in depth, more rounded and real, harder and more tangible. This result is due to slight differences between the two pictures, taken from two points a few inches apart. All the information to be revealed by the stereoscopic viewing is contained in these scarcely perceptible disparities. It should be possible to compute from them the spatial dimensions of the objects and their distribution in depth, and I could imagine cases in which the result of such processing may be of interest. But this would not tell us what the things photographed look like. If you want to remember a family party or identify a criminal, you must integrate the stereo-pictures by looking at them simultaneously with one eye on each.

When looking at the stereo-image, we do see the separate pictures too; for we see the stereo-image only because we have a precise impression of the two pictures which contribute to it. But we must distinguish between the two kinds of seeing: we are **focally aware**, while we are **subsidiarily aware**. We don't look at these two in themselves, but see them as clues to their joint appearance in the stereo-image. It is their function to serve as clues.

We may describe the situation by saying that we are **focally aware** of the stereo-image, by being **subsidiarily aware** of the two separate pictures. And we may add that the characteristic feature of subsidiary awareness is to have a **function**, the function of bearing on something at the focus of our attention. Next we may observe that the focal image, into which the two subsidiary pictures are fused, **brings out their joint meaning**; and thirdly, that this fusion **brings about a quality** not present in the appearance of the subsidiaries. We may recognize then these three features as parts of a process of knowing a focal object by attending subsidiarily to the clues that bear on it. We meet here the structure of **tacit knowing**, with its characteristic **functional**, **semantic**, and **phenomenal** aspects.

I have developed this analysis of tacit knowing many times before and have now chosen the example of
stereoscopic viewing in order to prevent a recurrent misconception. [2] It is a mistake to identify subsidiary awareness with subconscious or preconscious awareness, or with the fringe of consciousness described by William James. The relation of clues to that which they indicate is a *logical relation* similar to that which a premise has to the inferences drawn from it, but with the important difference that tacit inferences drawn from clues are not explicit. They are informal, tacit.

Remember that Helmholtz tried to interpret perception as a process of inference, but that this was rejected, because optical illusions are not destroyed by demonstrating their falsity. Tacit inference is like this. The fusion of the two stereoscopic pictures to a single spatial image is not the outcome of an argument; and if its result is illusory, as it will can well be, it will not be shaken by argument. The fusion of the clues to the image on which they bear is *not a deduction but an integration*.

Jean Piaget has drawn a striking distinction between a sensorimotor act and an explicit inference. Explicit inference is reversible: we can go back to its premises and go forward again to its conclusions, rehearse the whole process as often as we like. This is not true for the sensorimotor act: for example, once we have seen through a puzzle, we cannot return to an ignorance of its solution.

The seeing of two stereo-pictures as one spatial image is, indeed, irreversible in two senses. Firstly, it is difficult to find our way back to the clues in the two pictures, because they are hardly visible. And there are many other clues to seeing something, like memories and the feeling inside our eye muscles, which we either cannot trace or cannot experience in themselves at all; they are *largely submerged, unspecifiable*. Secondly--and this is more important for us--to go back to the premises of a tacit inference brings about its reversal. It is not to retrace our steps, but to efface them. Suppose we take out the stereo-pictures from the viewer and look at them with both eyes. All the effects of the integration are cancelled; the two pictures no longer function as clues, their joint meaning has vanished. What has happened here may be regarded as the inverse of tacit inference; a process of *logical disintegration has reduced a comprehensive entity to its relatively meaningless fragments*.

The best-known example of this is the way a spoken word loses its meaning if we repeat it a number of times, while carefully attending to the movement of our lips and tongue and to the sound we are making. All these elements are meaningful, so long as we attend through them to that on which they bear, but lose their meaning when we attend to them in themselves, focally. The famous tight-ropewalker, Blondin, says in his, memoirs that he would instantly lose his balance, if he thought directly of keeping it; he must force himself to think only of the way he would eventually descend from the rope.[3]

*The purpose of this paper is to show that the relation between body and mind has the same logical structure as the relation between clues and the image to which the clues are pointing.*

I believe that the paradoxes of the body-mind relation can be traced to this logical structure and their solution be found in the light of this interpretation.

The example of stereo-vision stands of course for a wide range of similar intellectual and practical feats of knowing. We know a comprehensive whole, for example a dog, by relying on our awareness of its parts for attending focally to the whole. When we perform a skill, we attend focally to its outcome, while being aware subsidiarily of the several moves we co-ordinate to this effect. I have carried out this analysis often elsewhere and shall take it for granted here.[4] But there is a further step which I must restate once more. I shall say that we observe external objects by being subsidiarily aware of the impact they make on our body and of the responses our body makes to them. All our conscious transactions with the world involve our subsidiary use of our body. And our body is the only aggregate of things of which we are aware almost exclusively in such a subsidiary manner.
I am speaking here of active consciousness, which excludes incoherent dreams or pathological bursts of temper. Active consciousness achieves coherence by integrating clues to the things on which they bear or integrating parts to the wholes they form. This brings forth the two levels of awareness: the lower one for the clues, the parts or other subsidiary elements and the higher one for the focally apprehended comprehensive entity to which these elements point. A deliberate act of consciousness has therefore not only an identifiable object as its focal point, but also a set of subsidiary roots which function as clues to its object or as parts of it.

This is the point at which our body is related to our mind. As our sense organs, our nerves and brain, our muscles and memories, serve to implement our conscious intention, our awareness of them enters subsidiarily into the comprehensive entity which forms the focus of our attention. A suitable term is needed to speak of this relation briefly. I shall say that we attend from the subsidiary particulars to their joint focus. Acts of consciousness are then not only conscious of something, but also conscious from certain things which include our body. When we examine a human body engaged in conscious action, we meet no traces of consciousness in its organs; and this can be understood now in the sense that subsidiary elements, like the bodily organs engaged in conscious action, lose their functional appearance when we cease to look from them at the focus on which they bear, and look instead at them, in themselves.

The way we know a comprehensive entity by relying on our awareness of its parts for attending to its whole is the way we are aware of our body for attending to an external event. We may say therefore that we know a comprehensive entity by interiorizing its parts or by making ourselves dwell in them; and the opposite process of switching attention to the parts can be described as turning the parts into external objects without functional meaning; it is to externalize them.

This formulation of tacit knowing, is particularly suited for describing the way we know another person's mind. We know a chess player's mind by dwelling in the stratagems of his games and know another man's pain by dwelling in his face distorted by suffering. And we may conclude that the opposite process, namely of insisting on looking at the parts of an observed behaviour as several objects, must make us lose sight of the mind in control of a person's behaviour.

But what should we think then of current schools of psychology which claim that they replace the study of mental processes, by observing the several particulars of behaviour as objects and by establishing experimentally the laws of their occurrence? We may doubt that the identification of the particulars is feasible since they will include many unspecifiable clues; but the feasibility of the programme will not only be uncertain, it will be logically impossible. To objectivize the parts of conscious behaviour must make us lose sight of the mind in control of a person's behaviour.

Admittedly, behaviourist studies do not reach this logical consequence of their programme. This is due to the fact that we cannot wholly shift our attention to the fragments of conscious behaviour. When we quote a subject's report on a mental experience in place of referring to this experience, this leaves our knowledge of that experience untouched; the report has in fact no meaning, except by bearing on this experience. An experimenter may speak of an electric shock as an objective fact, but he administers it only because he knows its painful effect. Afterwards he observes changes in the conductivity of the subject's skin which in themselves would be meaningless, for they actually signify the expectation of an electric shock--the skin response is in fact but a variant of goose flesh.

Thus a behaviourist analysis merely paraphrases mentalist descriptions in terms known to be symptoms of mental states and its meaning consists in its mentalist connotations. The practice of such paraphrasing might be harmless and sometimes even appropriate, but a preference for tangible terms of description will often be
restrictive and misleading. The behaviourist analysis of learning, for example, has banned the physiognomies of surprise, puzzlement, and concentrated attention, by which Koehler described the mental efforts of his chimpanzees. It avoids the complex, delicately graded situations which evoke these mental states. The study of learning is thus cut down to its crudest form known as conditioning. And this oversimple paradigm of learning may then be misdescribed as it was by Pavlov, when he identified eating with an expectation to be fed, because both of these induce the secretion of saliva. Wherever we define mental processes by objectivist circumlocutions, we are apt to stumble into such absurdities.

The actual working of behaviourism therefore confirms my conclusion that strictly isolated pieces of behaviour are meaningless fragments, not identifiable as parts of behaviour. Behaviourist psychology depends on covertly alluding to the mental states which it sets out to eliminate.

Principles of Boundary Control

But is not the material substance of all higher entities governed throughout by the laws of inanimate matter? Does it not follow then that it must be possible to represent all their workings in terms of these laws? Yes, this would follow. If I claim that these higher entities are irreducible, I must show that they are governed in part by principles beyond the scope of physics and chemistry. I shall do so. I shall show first that a number of different principles can control a comprehensive entity at different levels. I have repeatedly presented this theory in more particular terms.[5] It will be developed here on general lines.

There exist principles that apply to a variety of circumstances. They can be laws of nature, like the laws of mechanics, or be principles of operation, like those of physiology, as for example those controlling muscular contraction and co-ordination; or they can be principles laid down for the use of artifacts, like the vocabulary of the English language or the rules of chess. Not all important principles have such wide scope; but I need not go into this, for it is enough to have pointed out that some principles exist that do.

We can go on to note then that such a principle is necessarily compatible with any restriction we may choose to impose on the situation to which it is to apply; it leaves wide open the conditions under which it can be made to operate. Consequently, these conditions lie beyond the control of our principle and may be said to form its boundaries, or more precisely its boundary conditions. The term 'boundary conditions'--borrowed from physics--will be used here in this sense.

Next we recognize that in certain cases the boundary conditions of a principle are in fact subject to control by other principles. These I will call higher principles. Thus the boundary conditions of the laws of mechanics may be controlled by the operational principles which define a machine; the boundary conditions of muscular action may be controlled by a pattern of purposive behaviour, like that of going for a walk; the boundary conditions of a vocabulary are usually controlled by the rules of grammar, and the conditions left open by the rules of chess are controlled by the stratagems of the players. And so we find that machines, purposive actions, grammatical sentences, and games of chess, are all entities subject to dual control.

Such is the stratified structure of comprehensive entities. They embody a combination of two principles, a higher and a lower. Smash up a machine, utter words at random, or make chess moves without a purpose and the corresponding higher principle--that which constitutes the machine, that which makes words into sentences, and that which makes moves of chess into a game--will all vanish and the comprehensive entity which they controlled will cease to exist.

But the lower principles, the boundary conditions of which the now effaced higher principles had controlled,
remain in operation. The laws of mechanics, the vocabulary sanctioned by the dictionary, the rules of chess, they will all continue to apply as before. Hence no description of a comprehensive entity in the light of its lower principles can ever reveal the operation of its higher principles. *The higher principles which characterize a comprehensive entity cannot be defined in terms of the laws that apply to its parts in themselves.*

On the other hand, a machine does rely for its working on the laws of mechanics; a purposive motoric action, like going for a walk, relies on the operations of the muscular system which it directs, and so on. The operations of higher principles rely quite generally on the action of the laws governing lower levels.

Yet, since the laws of the lower level will go on operating, whether the higher principles continue to be in working order or not, the action of the lower laws may well disrupt the working of the higher principles and destroy the comprehensive entity controlled by them.

Such is the mechanism of a two-levelled comprehensive entity. Let me show now that the two-levelled logic of tacit knowing performs exactly what is needed for understanding this mechanism.

Tacit knowing integrates the particulars of a comprehensive entity and makes us see them forming the entity. This integration recognizes the higher principle at work on the boundary conditions left open by the lower principle, by mentally performing the workings of the higher principle. It thus materializes the *functional structure* of tacit knowing. It also makes it clear to us how the comprehensive entity works by revealing the meaning of its parts. We have here the *semantic aspect* of tacit knowing. And since a comprehensive entity is controlled as a whole by a higher principle than the one which controls its isolated parts, the entity will look different than an aggregate of its parts. Its higher principle will endow it with a stability and power appearing in its shape and motions and usually produce also additional novel features. We have here the *phenomenal aspect* of tacit knowing.

And finally, we are presented also with an ontological counterpart of the *logical disintegration* caused by switching our attention from the integrating centre of a comprehensive entity to its particulars. To turn our attention from the actions of the higher principle, which defines the two-levelled entity, and direct it to the lower principle controlling the isolated parts of the entity is to lose sight of the higher principle and indeed of the whole entity controlled by it. This mirrors the destruction of a comprehensive entity when it is pulled to pieces. The logical structure of tacit knowing thus covers in every detail the ontological structure of a combined pair of levels.

**Application of these Principles to Mind and Body**

The next question is whether the functioning of living beings and of their consciousness is in fact stratified. Is it subjected to the joint control of different principles working at consecutive levels?

The laws of physics and chemistry do not ascribe consciousness to any process controlled by them; the presence of consciousness proves, therefore, that other principles than those of inanimate matter participate in the conscious operations of living things.

There are two other fundamental principles of biology which are beyond the scope of physics and chemistry. The structure and functioning of an organism is determined, like that of a machine, by constructional and operational principles which control boundary conditions left open by physics and chemistry. We may call this a *structural principle*, lying beyond the realm of physics and chemistry. I have explained this a number of times before and will not argue it here again. [6]
Other functions of the organism not covered by physics and chemistry are exemplified by the working of the morphogenic field. Its principles are expressed most clearly by C. H. Waddington's 'epigenetic landscapes'. These show that the development of the embryo is controlled by the gradient of potential shapes, in the way the motion of a heavy body is controlled by the gradient of potential energy.[7] We may call this principle an organizing field or speak of it as an organismic principle.

Most biologists would declare that both the principles of structure and of organizing fields will be reduced one day to the laws of physics and chemistry. But I am unable to discover the grounds -- or even understand the meaning -- of such assurances, and hence I will disregard them and recognize these two principles as actually used in biology today.

Living beings consist in a hierarchy of levels, each level having its own structural and organismic principles. On the mental level, explicit inferences represent the operations of fixed mental structures, while in tacit knowing we meet the integrating powers of the mind. In all our conscious thoughts, these two modes mutually rely on each other, and it is plausible to assume that explicit mental operations are based on fixed neural networks, while tacit integrations are grounded mainly in organizing fields. I shall assume also that these two principles are interwoven in the body, as their counterparts are in thought.

The purpose of this paper is to explain the relation between body and mind as an instance of the relation between the subsidiary and the focal in tacit knowledge. The fact that any subsidiary element loses its meaning when we focus our attention on it explains the fact that, when examining the body in conscious action, we meet no traces of consciousness in its organs. We lose the meaning of the subsidiaries in their role of pointing to the focal. Using this principle, we are now ready to complete our project.

We have seen that we can know another person's mind by dwelling in his physiognomy and behaviour; we lose sight of his mind only when we focus our attention on these bodily workings and thus convert them into mere objects. But a neurophysiologist, observing the events that take place in the eyes and brain of a seeing man, would invariably fail to see in these neural events what the man himself sees by them. We must ask why the neurologist cannot dwell in these bodily events, as he could in the subject's physiognomy or intelligent behaviour.

We may notice that the latter kind of indwelling, for which we appear to be equipped by nature, enables us to read only tacit thoughts of another mind: thoughts and feelings of the kind that we may suitably ascribe to organismic processes in the nervous system. We can get to know the explicit thoughts of a person -- which may correspond to anatomically fixed functions of the nervous system -- only from the person's verbal utterances. The meaning of such utterances is artificial; though ultimately based on demonstrations pointing at tacit experiences, such utterances have no direct appeal to the native mind. The facility for indwelling can be seen to vary also when prehistoric sites, unperceived from the ground, are discerned from the air. Our incapacity for experiencing the neural processes of another person in the manner he experiences them himself may be aligned with these gradual variations of indwelling.

We arrive thus at the following outline. Our capacity for conducting and experiencing the conscious operations of our body, including that of our nervous system, lies in the fact that we dwell fully in them. No one but ourselves can dwell in our body directly and know fully all its conscious operations; but our consciousness can be experienced also by others to the extent to which they can dwell in the external workings of our mind from outside. They can do this fairly effectively for many tacit workings of our mind by dwelling in our physiognomy and behaviour, such powers of indwelling are fundamentally innate in us. By contrast, our explicit thoughts can be known to others only by dwelling in our pronouncements, the making and understanding of which is founded on artificial conventions. Objectivization, whether of another person's
gestures or of his utterances, cancels our dwelling in them, destroys their meaning and cuts off communication through them. The nervous system, as observed by the neurophysiologist, is always objectivized and can convey its meaning to the observer only indirectly, by pointing at a behaviour or at reports that we understand by indwelling.

The nervous system, as observed by the neurophysiologist, is always objectivized and can convey its meaning to the observer only indirectly, by pointing at a behaviour or at reports that we understand by indwelling.

The logic of tacit knowing and the ontological principles of stratified entities were derived here independently of each other, and we found that our tacit logic enables us to understand stratified entities. It shows us then that the higher principle of a stratified entity can be apprehended only by our dwelling in the boundary conditions of a lower principle on which the higher principle operates. Such indwelling is logically incompatible with fixing our attention on the laws governing the lower level. Applied to mind and body, as to two strata in which the higher principles of the mind rely for their operations on the lower principles of physiology, we arrive at three conclusions.

(1) No observations of physiology can make us apprehend the operations of the mind. Both the mechanisms and organismic processes of physiology, when observed as such, will always be found to work insentiently.

(2) At the same time, the operations of the mind will never be found to interfere with the principles of physiology, nor with the even lower principles of physics and chemistry on which they rely.

(3) But as the operations of the mind rely on the services of lower bodily principles, the mind can be disturbed by adverse changes in the body, or be offered new opportunities; by favourable changes of its bodily basis.

The way integration functions in tacit knowing, as well as the presence of irreducible organismic principles in living beings, are both consonant with the arguments, presented by Sir Francis Walshé for the presence of integrative, mental powers, not accounted for by the fixed anatomic structures of the central nervous system.[8]

Retrospect

Many philosophic efforts of our century can be see to have pointed towards such conclusions. A systematic attempt to safe-guard the content of unsophisticated experience against the effects of a destructive analysis was made by Edmund Husserl during the first three decades of this century with far-reaching influence on Continental philosophy. But its bearing on the body-mind problem was derived mainly later by Merleau-Ponty in his *Phénoménologie de la Perception* (1945). He gives a vivid and elaborate description of the way we experience our body. The body is 'known to us', he writes, 'through its functional value'; its parts engaged in the performance of our actions 'are available to us in virtue of their common meaning';[9] our body expresses meaning but 'language does not express thought, it is the subject's taking up of a position in the world of his meanings'.[10] 'If a being is conscious it must be nothing but a network of intentions;'[11] 'I do not understand the gestures of others by an act of intellectual interpretation.... The act by which I lend myself to the spectacles must be recognized as irreducible to anything else;'[12] our experience of our body is an existential act, not based either on observation nor on explicit thought. These remarks foreshadow my analysis, but I find among them neither the logic of tacit knowing nor the theory of ontological stratification, which I regard as indispensable for the understanding of the phenomena described by Merleau-Ponty.

Another follower of Husserl, Dr F. S. Rothschild, arrived even earlier at the conclusion that the mind is the meaning of the body. [13] He developed this idea widely in neurophysiology and psychiatry, where I am not competent to follow him.

The mainstream of contemporary English and American philosophy ignores the inquiries of
phenomenologists. But it shares their rejection of Cartesian dualism, and the kinship of the two movements goes beyond this. Deprive my quotations from the Phenomenology of Perception of their existentialist perspective, and they can be equated with observations of Ryle in the Concept of Mind (1949). But such a transition brings out the theoretical inadequacy of these observations and results in drawing false conclusions from them. Take a simple example. Merleau-Ponty says 'I do not understand the gestures of others by an act of intellectual interpretation', and Ryle says the same: 'I am not inferring to the workings of your mind, I am following them';[15] but Merleau-Ponty finds an alternative to 'intellectual interpretation' in existential experience, while Ryle has none and affirms, therefore, that 'most intelligent performances are not clues to the mind; they are those workings', [16] which is absurd. Many vivid and often subtle phenomenological descriptions are used by Ryle to demonstrate that the mind does not explicitly operate on the body, and from this result he concludes that body and mind are 'not two things', [17] 'not tandem operations', containing no, 'occult causes', [19] 'no occult antecedents' [20] no 'ghost in the machine', in other words, no Cartesian duality. But what actually follows from the fact that mind and body do not interact explicitly is that they interact according to the logic of tacit knowing. And it is this logic that disposes of the Cartesian dilemma by acknowledging two mutually exclusive ways of being aware of our body.

As Ryle's powerful argument leads him to fallacious conclusions, it offers a compelling demonstration of the troubles arising from the absence of the cognitive and ontological principles outlined in the present paper; that is why I selected his work for representing anti-Cartesian thought in contemporary British and American literature.

Notes


[13] See Rothschild's earlier writings, which extend back to 1930. A fairly recent summary of them is given in the monograph: F. S. Rothschild, *Das Zentralnervensystem als Symbol des Erlebens*, Basel and New York: S. Karger, 1958, VII, pp. 1-134. In this monograph, Dr. Rothschild points out on pp.10-11 that the meaning of the CNS manifested in consciousness is lost by examining the CNS as an object - just as the denotative meaning of a word is lost by such an examination. This anticipates part of my theory of body and mind. For a briefer summary in English, of Dr. Rothschild's work, see F. S. Rothschild, 'Laws of Symbolic Media-tion in the Dynamics of Self and Personality', *Annals of The New York Academy of Sciences*, 96 (1962), pp. 774-84.


