Australasian Journal of Philosophy
Vol. 46, No. 2; August, 1968

NICHOLAS MAXWELL

UNDERSTANDING SENSATIONS

I

My aim in this paper is to defend a version of the brain process theory, or identity thesis, which differs in one important respect from the theory put forward by Professor Smart.[1] I shall argue that although the sensations which a person experiences are, as a matter of contingent fact, brain processes, nonetheless there are facts about sensations which cannot be described or understood in terms of any physical theory. These 'mental' facts cannot be described by physics for the simple reason that physical descriptions are designed specifically to avoid mentioning such facts. Thus in giving a physical explanation of a sensation we necessarily describe and render intelligible that sensation only as a *physical process,* and not also as a *sensation.* If we are to describe and render intelligible a person's sensations, or inner experiences, as *sensations,* and not as physical processes occurring in that person's brain, then we must employ a kind of description that cannot be derived from any set of physical statements.

The kernel of the argument of this paper may be expressed as follows. There are neurophysiological processes which can be understood as *sensations,* as opposed to physical processes, only if sufficiently similar neurophysiological processes have occurred in one's own brain. More precisely, there are facts about certain neurophysiological processes which are such that there can be no description of these facts whose *meaning* one can understand unless sufficiently similar neurophysiological processes have occurred in one's own brain. But a person who has not had these neurophysiological processes occur in his brain is not thereby debarred from completely understanding a complete *physical* description of such neurophysiological processes. It follows that a complete physical description of these neurophysiological processes, supposing such a thing were possible, would not be a *complete* description: it would not tell us all that there is to know about the processes in question.

It might be thought that since the version of the identity thesis defended here must imply that certain brain processes have mental features, in some sense of 'mental', in addition to their ordinary physical features, this

[1] See J. J. C. Smart, *Philosophy and Scientific Realism,* Routledge and Kegan Paul, 1963, Ch. 5.

version of the thesis must reintroduce just those conceptual difficulties which Professor Smart's version avoids. I shall argue however that there is nothing unintelligible or inexplicable about the mental feature of a brain process, as construed here, and no *conceptual* problem as to how the mental and physical features are inter-related. I hope to show that the nature of sensations, or inner experiences, may *seem* unintelligible or inexplicable precisely because the false assumption is made that if sensations are to be understood at all then it must be possible to describe and understand them in terms of some physical theory. Again, it may be thought that there is .a conceptual problem about the relation between the mind and the brain just because the above dualism of description (physical and non-physical) is neglected. The genuine, but unacknowledged, dualism of description may delude us into believing in a dualism of substance, mind and brain.

II

The brain process theory, or identity thesis, as formulated by U. T. Place,[2] Professor Smart and others, states that all sensations or inner experiences are brain processes. The theory implies not that the two phrases 'such and such inner experience' and 'such and such brain process' ever have the same *meaning,* but only that the first phrase, as a matter of contingent fact, refers always to some brain process. Thus the theory is put forward as an extremely plausible empirical hypothesis; it is not intended to provide an analysis of 'sensation' or 'inner experience'.

The theory presupposes that:

(1) In experiencing a sensation a person is presented with evidence for the existence of some entity or process, and is not merely disposed to behave in a certain way.

(2) A person who experiences sensations on two distinct occasions is, in general, in a position to judge whether or not the entities or processes, for whose existence he has evidence, are similar or dissimilar with respect to some feature or other. (This leaves open the question as to whether or not the person is in a position to know in what respect the two entities or processes are similar or dissimilar.)

In what follows I shall use the word 'sensation' to refer to that entity or process, whatever it may be, for whose existence one has evidence in `experiencing a sensation' or 'having an inner experience', and I shall say two sensations are similar if the person who is presented with evidence for the existence of the sensations, judges them to be similar. The two phrases 'experience a sensation' and 'have an inner experience', I shall use interchangeably, with no technical meaning assigned to them. The brain process theory may now be stated as follows:

---

[2]U. T. Place, 'Is Consciousness a Brain Process', *British Journal of Psychology,* 1956, 47, pp. 44-50.

(3) Sensations are, as a matter of contingent fact, brain processes This means at least (a) the distinctive features of sensations and brain processes may, without conceptual difficulties, be ascribed to one and the same thing; (b) two sensations are similar if and only if similar physical processes occur in the brain of the person in question.

In addition to (3) it may be held that:

(4) All features or characteristics of sensations can in principle be described, explained, understood physically.

I wish to argue that (3) is a plausible empirical hypothesis, fully in accordance with ordinary experience and present day knowledge, but that if (3) is accepted, (4) must be rejected. I shall give only an outline of the first part of this argument, since here I am in essentials in agreement with Professor Smart.

In what follows I shall consider only *visual* sensations, but similar remarks will apply to the other kinds of sensations.

I assume here without discussion the truth of (1), (2) and (3b). Granted this, it will I think be conceded that in order to establish that (3) is a plausible empirical hypothesis it is necessary and sufficient for me to establish that:

(5) All that a person knows about that entity or process, for whose existence he has evidence in having an inner experience, is compatible with the hypothesis that the entity is a brain process.

Clearly the crucial point to be considered here is this: what precisely does a person know in experiencing a certain visual sensation? Professor Smart considers the case of a person who experiences a yellowish-orange after-image, and asserts, in effect, that such a person knows only *'What is going on in me is like what is going on in me when* my eyes are open, the lighting is normal, etc., etc., and there really is a yellowish-orange patch on the wall'.[3] Professor Smart seems to take this to imply that the person knows only that some process is occurring which is similar, in some wholly unknown respect, to some other process.[4] This conclusion I wish to contest. I maintain that a person who experiences the above after-image knows not merely 'A process is occurring which is similar, in some *wholly* unknown respect, to some other process', but also 'Whatever it is that is going on in me, it is such that if certain other conditions—necessary but not sufficient conditions for perception—had been fulfilled (e.g. if there really had been a yellowish-orange patch on the wall, my eyes had been open, etc., etc.), then I would have been perceiving a yellowish-orange patch'. We might say, condensing all this, that a person who experiences a yellowish-orange after-image knows 'It is, for me, at this moment, just as if I am perceiving a yellowish-orange patch', or 'My state is such that it

[3] J. J. C. Smart, *op. cit.*, p. 94.

[4] See *ibid.*, p. 95.

is just as if I am perceiving a yellowish-orange patch'. As long as 'This patch is yellowish-orange' can constitute a true description of an actual patch, 'My state is such that it is just as if I am perceiving a yellowish-orange patch' can be interpreted as constituting a genuinely informative description of the state in question. It is because the person possesses this slender additional item of knowledge about what is going on in him that he is able to classify or describe what is going on in him as a visual sensation of a *yellowish-orange patch.*

It is important to note however that this additional knowledge about what is going on in him does not conflict with the hypothesis that what is going on is some brain process. It is, in Professor Smart's terminology, `topic-neutral'. Since a person who experiences a visual sensation knows this much and no more about what is going on in him, this establishes (5), and hence the plausibility of (3).

I might add that my position here does not I think differ substantially from Professor Smart's. It is only when the question arises 'What meaning can be assigned to 'yellowish-orange' such that the proposition Tor some x, x is yellowish-orange' is true?' that our respective views really diverge.

A consequence of the above argument, vital for the argument of section IV below, is that a second person can understand that which the first person knows and understands in experiencing the yellowish-orange after-image only if he can understand the meaning of the term 'yellowish-orange'. `It is for me just as if I am perceiving a yellowish-orange patch' can only be understood if 'yellowish-orange' is understood.

It should perhaps be pointed out that a person may experience a visual sensation and not be in a position to know even this much about what is going on in him A person who has just been cured of congenital blindness and who experiences a visual sensation that he would later describe as the visual sensation of a yellowish-orange patch, probably cannot be said to know at the time that this is what he is experiencing, in that he does not yet fully possess the concept, fully understand the meaning of, `yellowish-orange'. In order fully to understand the meaning of colour words it is perhaps necessary to be able to perceive coloured objects, or discriminate between objects with respect to their colour, and this the above person may not initially be able to do (although presumably such a person *would* be able to discriminate between having a red light, say, and a blue light shone into his eyes). I shall discuss later (see section V) the question: Does 'yellowish-orange' have some kind of minimal meaning which *could* be understood by the person who has just been cured of congenital blindness? Here, I assume (in accordance with (1) above) that a person can judge whether or not two successive visual sensations he experiences are similar, even if he does not fully possess the relevant perceptual concepts.

It may be objected that there is something very puzzling about this assumption, in that, in view of the discussion above, it appears to imply that a person can judge whether or not two processes are similar even if he knows nothing about any respect in which the two processes are similar. It should be noted that Professor Smart accepts this implication.' The argument of this paper, however, does not presuppose this Smartian thesis. In section V, I shall argue that a legitimate meaning *can* be assigned to `yellowish-orange' which could be understood by the person who has just been cured of congenital blindness, and hence that such a person is in a position to know something about the sensations he experiences (e.g. 'I am experiencing yellowish-orange' in the above sense of `yellowish-orange'), even though he does not fully possess the perceptual concept 'yellowish-orange'.

III

Before giving my reasons for maintaining that in experiencing a sensation a person is in a position to know something about that which is happening which cannot be described or understood in terms of any physical theory, it will be convenient to consider first the following problem: what interpretation can one reasonably give to the statement `Two persons are experiencing similar sensations', granted the truth of the above brain process theory?

Three points should be noted.

(a) The problem of giving an interpretation to 'Two persons are experiencing similar sensations' is not peculiar to the above brain process theory, since it is not clear what we should ordinarily want to mean by this statement, irrespective of whether or not we have accepted the above theory.

(b) The meaning of 'Two persons are experiencing similar sensations' is problematic because it is not clear how even in principle the sensation of one person could be compared with the sensation of another. In order to give an unproblematic interpretation to the above sentence it will suffice to specify how in principle a sensation experienced by one person can be compared with a sensation experienced by another.

(c) Our ordinary thoughts and talk about sensations tend to be infected with Cartesian dualism, with the view that sensations are peculiarly `mental' entities, wholly distinct from brain processes. Hence we must not be surprised if an unproblematic interpretation of 'Two people experience similar sensations', which is compatible with the above brain process theory, fails to capture all that we might hope to mean by this sentence. A shift in ordinary meaning is bound to occur when a theory, whose truth is presupposed implicitly in the meaning of certain words, turns out to be false.

[5] See *ibid.*, p. 95-6.

The problem, then, is to determine how one person, X, could in principle discover what some other person, Y, is experiencing, so that he can compare Y's sensation with his own. But, one wants to say, Y's sensations or inner experiences are essentially private; only if X could achieve the impossible and actually become Y, could X discover what Y is experiencing.

X cannot actually *become* Y, but X *can*, in principle, (we may assume) become something which is in all relevant physical respects, very similar to Y; and this we may regard as being sufficient for X to discover what Y is experiencing. I suggest, in other words, that the meaning of 'X and Y are experiencing similar sensations' is to be interpreted in such a way that, given that:

(a) a brain process A is occurring in X's brain, a brain process B in Y's;

(b) X experiences sensations similar to the above if and only if A occurs in his brain (and similarly for Y visa vis B);

(c) A and B occur successively in the brain of any one person, Z;

then *Z's comparison of the sensations he experiences constitutes a comparison of the sensations experienced by X and* Y.

Given this 'brain process' interpretation, the meaning of 'X and Y are experiencing similar sensations' is unproblematic, in that it is in principle possible for anyone to compare X's and Y's sensations.

This interpretation of the above sentence seems to me to be reasonably close to what one would ordinarily want this sentence to mean. If Z is to discover what X is experiencing then surely, on any view, Z must discover what it would be like to have happen to himself that which is happening to X—which, granted the above brain process theory, is reasonably interpreted as: If Z is to discover what X is experiencing then Z must reproduce in his own brain relevant physical processes occurring in X's brain. Only if sensations are thought of as entities or processes entirely distinct from anything going on in the brain will the above interpretation of 'X and Y are experiencing similar sensations' seem far removed from anything that we should ordinarily want to mean by this sentence. But there is, it seems, no evidence whatsoever for the existence of such entities or processes. And in any case, once one conceives of sensations as peculiarly mental entities, entirely distinct from brain processes, it becomes wholly obscure what could possibly be meant by 'Two persons are experiencing similar sensations', since there is no conceivable way in which such mental entities existing in different minds could be compared.

As far as I can see, the only coherent alternative to the above brain process interpretation, is an interpretation which presupposes that two people who can make the same discriminations with respect to some kind of physical stimulus, experience similar sensations when exposed to the same stimulus. But the above brain process interpretation has one

immense advantage over this 'behavioural, discriminatory-response' interpretation. We should ordinarily want to say that it is possible for two persons to make identical discriminations with respect to some particular kind of stimulus (e.g. visible light of various wavelengths), and yet to have quite different sensations when exposed to precisely the same stimulus (e.g. light of a particular wavelength). If we adopt the behavioural interpretation, this is not a conceptual possibility. But if we adopt the brain process interpretation, this is quite straightforwardly possible, since the two persons may make identical discriminations, and yet sufficiently dissimilar physical processes may occur in their respective brains when they are exposed to precisely the same stimulus. Thus not only is the brain process interpretation of the sentence 'Two persons are experiencing similar sensations' reasonably close to what one would ordinarily want to mean by this sentence (which is all that was required), but further this interpretation is closer to the ordinary meaning than the only apparent alternative, the behavioural interpretation.

One or two detailed points need to be made about the above brain process interpretation. In the first place the interpretation is clearly only possible if any two persons who undertake to compare the sensations of X and Y in the above manner, obtain the same result. Thus the interpretation presupposes that if one person, X, has two different processes A and B occur successively in his brain and experiences similar sensations, then if any other person, Y, has A and B occur successively in his brain, Y also will experience similar sensations. This may quite reasonably be regarded as a (potentially) falsifiable, and therefore empirical hypothesis, a crucial point being that X (or anyone else) can test whether Y has, on the two occasions, similar sensations, in that he can test whether Y can differentiate between A and B occurring in his (i.e., Y's) brain.

The interpretation does not imply that if Z is to compare the sensations of X and Y he must duplicate precisely the brain states of X and Y in his own brain. Instead Z can determine to what extent the brain states of X and Y can be varied without X and Y being able to make any corresponding discriminations. In this way it will be possible for Z to determine a particular type of brain process that corresponds for X, or for Y, to a particular type of sensation.

One questionable assumption has been made above, namely that any type of brain process that occurs in the brain of one person can be reproduced in the brain of any other person. But this in general may not be possible: the structure or material of the brain of one person, X, may differ from that of the brain of another person, Y, to such an extent that many physical processes that occur in X's brain cannot occur in Y's, and vice versa. One might attempt to overcome this difficulty by prescribing for X, say, brain surgery sufficiently drastic to permit the occurrence in his brain of all processes that occur in Y's brain. But at the most this would

ensure only that *some* person could compare X's and Y's sensations, not that X could do this, since we might not want to say that X, as an individual, would survive the operation. On the other hand perhaps this is sufficient for our purposes. Thus we may stipulate that it is meaningful to assert that X and Y are experiencing similar, or dissimilar, sensations, as long as it is possible for there to be some person, Z, whose brain is such that the physical processes occurring in X's brain and in Y's, can occur in it without bringing about the destruction of Z. However in what follows I shall make the empirical assumption that such problems would arise only if human beings were to try to ascertain whether or not such things as Robots and Martians experience sensations similar to those experienced by human beings (the Robots and Martians in question *behaving* just as if they were human beings). **I** assume, in other words, that the structure and material of any two human brains are sufficiently similar to permit relevant processes that occur in one brain, to occur in the other brain.

IV

I come now to the main part of my argument, my defence of the thesis that it is impossible to describe or understand physically all that a person can know in experiencing a sensation.

Consider any particular property or quality Q which does actually exist in the world, i.e., which can be truly ascribed to some objects. In order to become aware of the nature of Q, i.e., in order to understand the meaning of the term 'Q', it will be necessary to have some experiences, and therefore, according to (3), necessary to have some brain processes occur in one's own brain. Now there are two possibilities. On the one hand it may not be necessary to experience any particular kind of sensation. (Two sensations are of the same kind if, roughly, they are sufficiently similar, where similarity of sensations is interpreted as above.) But on the other hand it may be that the meaning of '0' can only be understood if one has experienced a particular kind of sensation, and hence, according to (3), if one has had a particular kind of brain process occur in one's own brain. This possibility cannot be excluded on a *priori* grounds. Obvious candidates for such properties are colours, sounds, smells and tastes as ordinarily conceived. For example, redness is such a property if (a) there are some red objects, (b) a person who has not experienced the visual sensation of redness, e.g., someone who **is** congenitally blind, cannot fully understand the meaning of the word 'red'.

Assuming for the moment that a visual property **P** of this latter kind does exist, suppose one person, X, is experiencing the visual sensation of **P,** i.e. a certain kind of brain process, A say, is occurring in his brain. Suppose further that X frequently perceives objects with the property **P,** and understands the meaning of **'P'.** Now assume that there is a second

person, Y, who has never had A occur in his brain, who cannot, therefore, understand the meaning of 'P', and who, further, cannot perceive that objects have the property P, i.e., whenever he looks at an object that does have P, A does occur in his brain. (If `P' is 'redness' then Y might be congenitally blind.) My argument is that Y *is not thereby prevented from giving a complete physical description and explanation of all that happens to X and all that X does;* he may for example be able to predict and explain the occurrences of A in X's brain, and the perceptual discriminations with respect to P that X is able to make. Nonetheless there is one fact about what is happening to X, which X knows and understands, but which must remain completely unintelligible to Y, namely that X is experiencing the visual sensation of P, i.e. that it is for X just as if he is perceiving an object with the property P. Y cannot understand this because he cannot understand the meaning of the term 'P'. By hypothesis, this fact becomes intelligible to Y only if A occurs in his brain. Hence, 'X is experiencing the visual sensation of P' is a fact about what is happening to X which cannot be described or understood in terms of any physical theory.

During the course of this argument I have assumed that:

(6) There are what may be called P-properties, that is, properties of the above type, which (a) are perceived by human beings, (b) are not *physical* properties, i.e. cannot be referred to by any purely *physical* term or statement.

I shall now attempt to justify this assumption.

As far as I am concerned, (6a) scarcely seems to be an assumption that requires any extended philosophical argument for its justification: my most casual observation appears to verify conclusively that properties such as colours, sounds, smells, tastes exist which I could have had no conception of if I had not myself experienced the appropriate kinds of sensations. In order to understand all that I would wish to mean by the word 'red', for example, it is not sufficient to be able to discriminate between objects that I perceive to be red, and objects that I perceive to have some other colour: a congenitally blind person might possess a piece of apparatus which enabled him to do this, and yet it seems clear that such a person could not fully understand what I mean when I assert of an object that it is red. In order to be able to understand this it is necessary to have experienced a sensation sufficiently similar to a sensation that I would describe as the visual sensation of a red patch.

The word 'red' can of course be assigned a meaning, a 'behavioural' or `discriminatory-response' or 'physicalistic' meaning (`red_s' say), such that the above congenitally blind person would be able to understand 'This is red_s'. Further, since it is possible that two persons, X and Y, who make precisely the same colour discriminations, may nonetheless experience dissimilar sensations when looking at some object they both describe as 'red' (where 'dissimilar sensations' is interpreted as above), it is possible that

the 'public' meaning of 'red' is equivalent merely to the meaning of 'red,', i.e. that which the above congenitally blind person can understand. It is possible, in other words, that in asserting of some object that it is red we manage to communicate to some non-colour-blind person no more than that which the above congenitally blind person can understand. Nevertheless the word 'red' can legitimately be assigned a meaning (`red_2' say), by me, for example, such that 'red,' can be understood only if one has experienced a certain kind of visual sensation. There is nothing *in principle* 'private' about the meaning of 'red_2', since, as I have argued above, it can in principle be determined whether or not two persons experience similar or dissimilar sensations. If my ordinary experience does not confirm the truth of 'For some x, x is red_2', then I find it difficult to see how any statement whatsoever can be said to be confirmed by means of experience.

I assume the reader is presented with similar convincing perceptual evidence for the existence of P-properties. Clearly, as long as `P-properties exist' does not conflict with any empirical statement that we are inclined, for one reason or another, to believe to be true, we have every reason to believe, and no reason to disbelieve, that P-properties such as colours, sounds, smells and tastes do exist.

Assuming, for the moment, that there is no such empirical statement which conflicts with the thesis P-properties exist', I turn now to my defence of (6b).

One might attempt to justify (6b) as follows. A physical statement (description, explanation or theory) is by definition an *objective* statement. But 'objective statement' means just 'statement that is in principle intelligible, meaningful, to any rational being'.[6] Now suppose there is a property which is such that the meaning of any term, `P' say, which refers to the property, can be understood only if the brain process A has occurred in one's own brain. Presumably it is in principle possible for a being to be rational even though the structure or material of the being's brain is such that the physical process A cannot occur in it. It follows that any statement which ascribes P to some object cannot be objective, in the above sense, and hence cannot be a *physical* statement.

This argument breaks down however if the assumption that physics is, by definition, objective in the above sense, is rejected. I give below therefore an alternative defence of (6b), which does not depend on the assumption that any physical theory or explanation is, by definition, objective in the above sense. I begin by defending the thesis that at least `red_2' (interpreted as above) cannot be a physical term. I have condensed

[6] For such an explication of 'objective' see my paper 'Physics and Common Sense', *British Journal for the Philosophy of Science*, Vol. XVI, No. 64, 1966, pp. 310-1.

my defence of this thesis into the following four remarks (bolstered up by arguments where necessary).

1. The task of physics is to predict and explain natural phenomena, that is, to predict and explain the outcome or result of any possible physical experiment. Any term whose meaning is such that it cannot be required for the physical prediction or explanation of some possible physical experiment is not a physical term.
2. There is no term, statement, theory of contemporary physics that refers to the property red2.

The terms of physics have an agreed, 'public' meaning; but since at the moment it is not possible to compare the sensations experienced by different people when they look at objects, or rays of light, that they classify as 'red', we have no reason to suppose that the agreed, 'public' meaning of 'red' is `red,'. If 'red' qualifies as an observational term of physics (and this may well be denied), then 'red' must be interpreted as `red_s' and not `red,'. In other words in order to understanding the meaning of 'red', construed as a *physical* term, it is, conceivably, necessary to be able to discriminate between those objects ordinarily classified as 'red', and those objects ordinarily classified as `non-red'; but it is not necessary in addition to experience a certain kind of sensation when looking at an object generally classified as 'red'. Hence 'red', construed as a physical term, cannot be interpreted as `red,', as referring to the P-property red_2.

It should be noted that, since physics cannot, at least at the present time, be required to predict the existence of red_2, the fact that no current physical theory refers to, or predicts the existence of red_2, constitutes *no evidence whatsoever* for the thesis that red_2 does not exist. The fact that physics today makes no mention of red_2 is evidence only for the thesis that the physicist does not need to refer to red_2 in order to predict and explain the kind of experimental results at present under review.

3. As long as physics is not required to predict the existence of the property red,, that is, as long as a statement of the form 'x is red_2' or 'X is experiencing the visual sensation of red_2' cannot qualify as a statement of the result of a physical experiment, no term that refers to the property red_2 can be required for the physical prediction or explanation of the result of any physical experiment.

If current physical theories are taken into account, it is almost impossible to conceive how the above thesis could be false. Certainly there is no known physical experiment, granted the above restriction, which is such that it is possible to conceive how the physical explanation of the result of the experiment could require the inclusion of a term that referred to red_2. It seems very improbable indeed that such an experiment could ever crop up. If such an experiment *is* possible, then our present physical theories must be very seriously astray indeed. Of course all present day

theories may turn out to be false; but not, one is inclined to say, that *false*.

It might be asked: But suppose one wants to give a physical explanation of either (a) a person's ability to discriminate between objects ordinarily classified as 'red' and objects ordinarily classified as `non-red'; or (b) the processes that occur in a person's brain when he looks at objects ordinarily classified as 'red'. Can one be certain that an explanation that incorporated the term 'red,', let us say, could not qualify as a *physical* explanation?

It must be admitted that it is not at present possible to give a completely satisfactory physical explanation of (a), let alone (b). But this is due to the great complexity of the eye and the brain. We have no reason to believe that current physical theories are, in themselves, inadequate for the solution of this problem; rather it is our ability to apply these theories to this particular problem that is inadequate. But in any case it is clear that a satisfactory solution to the above problem,, far from incorporating the term 'red,', would not even incorporate the term 'red,' (but instead terms such as 'light of such and such wavelength', that is, terms whose meaning can be understood even though the meaning of 'red,' or 'red,' is not understood). Any explanation that incorporated 'red,' or `red,' in an essential way would be completely unsatisfactory as a physical explanation because it would be wholly *ad hoc:* the explanation would incorporate a theory which could be applied *only* to the highly specific phenomenon (a) or (b). For example, one might attempt to give a physical explanation of (b) in terms of some such theory as: 'If an object 0 is red, and if a physical system, S, of such and such a type, is related to 0 in such and such a way, then the process A occurs in the 'brain' of S'. (Here, S is such that the body of the person in question is an S-type system.) But such an 'explanation' would be completely *ad hoc:* it could not possibly qualify as a physical explanation of the occurrence of A at all. For the above 'theory' could be applied only in the extremely complex, arbitrarily restricted circumstances that the 'theory' specifies; the 'theory' could not possibly be derived from some more comprehensive physical theory that applied to a wide range of phenomena, because no set of physical statements (which do not incorporate the term 'red,') could entail a 'theory' of the above type that does incorporate the term 'red,'.

4. Physics is not required to predict the existence of red,, that is, no statement of the form `x is red,' or 'X is experiencing the visual sensation of red,' can qualify as a statement of the outcome or result of a physical experiment.

Any physical experiment can I think be stipulated to be such that the outcome or result of the experiment can in principle be recorded in an objective, 'public' manner: clearly if 'objective' is interpreted as above,

no statement of the result of a physical experiment can refer to a P-property. More specifically, any physical experiment can be stipulated to be such that the outcome or result of the experiment can in principle be recorded instrumentally, as a configuration of objects or marks, e.g., as the position of a pointer on a dial, or as lines, dots or areas of exposure on a photographic plate. There would be something very odd indeed about a physical experiment which was such that its result could not in principle be recorded in this kind of way. In other words, without loss of generality, it can be stipulated that a necessary condition for an experiment to be a *physical* experiment is that the statement of the outcome or result of the experiment is such that in order to understand the meaning of this statement, it is not necessary to have experienced a particular kind of sensation. This is not to deny of course that if a person is going to verify such a statement of the outcome of an experiment, that person must of necessity experience sensations of some kind or other, e.g. visual or tactile. My point is only that if an experiment is to qualify as a physical experiment then the statement of the outcome of the experiment (whose verification involves experiencing such and such visual sensations say for one person, such and such tactile sensations for another person) must be such that in order to understand the meaning of this statement, it is not necessary to have experienced a particular kind of sensation.

In adopting the above necessary condition for an experiment to be a *physical* experiment one is in effect imposing a restriction on what one chooses to mean by the term 'physical'. It must be admitted that, in a sense, one is at liberty to assign whatever meaning one pleases to any term; in particular one might assign a meaning to 'physical' which is such that any singular existential statement qualifies as a possible statement of the result of a physical experiment, in which case, of course, it would be a contradiction in terms to assert that non-physical P-properties exist. Here I wish to say only that the above restriction that I have imposed on the meaning of 'physical' is non-arbitrary, and is in accordance with what seems to be ordinarily meant by 'physical'. (It might be noted however that even if this restriction is rejected, one would still not be able to give a satisfactory physical explanation of the existence of a P-property such as red,, for the 'explanation' would involve *ad hoc* theories somewhat similar to the *ad hoc* theory discussed under remark 3 above.)

One might argue, in a somewhat Smartian frame of mind, that the ultimate aim of physics is to discover, or formulate, a *comprehensive* physical theory, which, ideally, would postulate and describe just a few different kinds of 'fundamental physical entities'. Granted that there are no valid general or philosophical objections to interpreting physical theories realistically, the above kind of theory could be interpreted as asserting that the world is, in a sense, made up entirely of the fundamental physical entities in question. How can this be reconciled with the thesis that

physical theories are not required to predict the existence of P-properties, granted that such properties exist?

The answer to this is simply that the meaning of `comprehensive' in this context is such that a *comprehensive* physical theory suffices in principle to predict and explain the result of any possible *physical* experiment. In asserting that the world is made up entirely of fundamental physical entities, one is asserting that, given any possible physical experiment, the so-called initial conditions and the outcome of the experiment can in principle be specified or described in terms of the fundamental physical theory, and hence that this theory is comprehensive, in the above sense.

As long as necessary and sufficient physical conditions can in principle be specified for the existence of a P-property such as red_2, the physicist is entitled to claim that his neglect of this feature of things does not mean that there are *entities* to which physical theories do not apply.

I conclude from the above discussion that `red_2' cannot qualify as a physical term, and, in general, that such P-properties as colours, sounds, smells and tastes cannot be referred to by any physical term or statement.

So far I have assumed without discussion that there are no objections to the thesis that P-properties exist. Professor Smart has however put forward certain arguments which purport to establish the extreme implausibility of the thesis that P-properties exist.? But these arguments hinge on the contention that the thesis that P-properties exist (or, in Professor 'Smart's own terminology, the thesis that objective, unanalysable, intrinsic *quale* exist) is incompatible with the main body of scientific knowledge. And this contention is, I have argued, mistaken. The fact that physical theories do not refer to or predict the existence of P-properties constitutes no evidence whatsoever for the thesis that such properties do not exist.[8]

Thus we have every reason to believe, and no reason to disbelieve, that such non-physical P-properties as colours, sounds, smells and tastes do exist.

One might say, speaking rather loosely, that a physicist who could treat all human beings merely as complicated pieces of physical apparatus, distinct from himself, would have no reason to suspect the existence of P-properties. He would not require the hypothesis that such properties exist in order to predict and explain the workings and behaviour of his pieces of apparatus. But the human physicist cannot treat *all* human beings in this way, for he is *himself* such a piece of apparatus; consequently he cannot help but perceive colours, sounds, smells, etc. In so far as one is a physicist one assumes that one learns only by observing and theorizing about the behaviour of pieces of apparatus; one neglects

---

J. J. C. Smart, *op. cit.*, pp. 66-75.

[6] For a more detailed criticism of Professor Smart on this point, see my paper `Physics and Common Sense', *op. cit.*, pp. 295-311.

that one is oneself a piece of apparatus of a certain kind, and that there are, as a result, things which one has discovered which could have been discovered in no other way.

V

This in a sense concludes my defence of the thesis that although sensations are brain processes, nonetheless there are facts about sensations which cannot be described or understood in terms of any physical theory. There is however one question that **I** have deliberately, for expositional reasons, left unanswered. Suppose one person, X, experiences a sensation similar, let us say, to that which some other person, Y, would describe as the visual sensation of a red patch. Suppose further that X does not fully understand the meaning which Y gives to 'red', perhaps because X has only just been cured of congenital blindness, and hence is not in a position to know 'It is for me just as if **I** am perceiving a red patch'. The question is: Does X know anything about that entity, for whose existence he has evidence in experiencing the above sensation? This question can be reformulated as follows: Can a sort of minimal sensory meaning be given to 'red' (`red_3' say) such that '**I** am experiencing red.,' can be true and meaningful to X?

It would **I** think be possible to defend a version of the above brain process theory, T, say, which gives a negative answer to this question. But to my mind a version, T_2 say, which gives an affirmative answer to this question is much more plausible. For surely a person who experiences sensations that we would judge to be visual sensations of red, green and blue patches, does know *something* about what is going on in him, even if he does not fully possess the perceptual concepts 'red', 'green', 'blue'.

Suppose a certain kind of brain process, A, can be induced artificially in the brain of a person, X, who is congenitally blind. Suppose a second person, Y, frequently perceives red objects (i.e., understands `red,'), knows that A occurs in his brain when and only when he experiences sensations of redness, and knows further when A occurs in X's brain. Now surely Y can teach X to use the word `red,', so that X can later convey information to Y (or indeed to a second congenitally blind person, X', who has also been taught by Y to use `red,') by asserting '**I** am experiencing $red_3$'. Note that (a) X may not be able to understand fully the meaning of 'red,' or `red2', (b) a fourth person, Z, who can understand a complete physical description of A, but who has not had A occur in his own brain, cannot understand what X, X', and Y understand by '**I am** experiencing red_3'.

That which X and Y understand and Z does not understand, one might describe (but *not* analyse) as 'What it is to have A occur in one's own brain'. This tells Z (a) why he does not understand that which X and Y understand, even though he can give a complete physical description of A,

(b) what he must do in order to be able to understand that which X and Y understand.

I conclude that T_2 is to be accepted and T_1 rejected.

At first sight it might seem that T_2 reintroduces just those conceptual difficulties concerning the relation between 'mental' and physical features that T_1 successfully avoids. Certainly there is a sense in which T_2 permits one to know a little more about one's sensations than that which T_i permits. But the important point to note is that this little additional knowledge does not constitute additional knowledge about the physical or perceptual properties that the sensation either does or does not possess. Hence if T_1 successfully avoids the above kind of conceptual difficulties, so does T_2.

It should be noted however that as far as the main argument of this paper is concerned (i.e. the argument of section IV), it is quite irrelevant whether one decides to accept T_2 or T_1.

VI

Both T_2 and T_1 can be interpreted as attributing 'mental' features to the events that occur within our heads, in addition to the ordinary physical features of such events. Here, 'mental' feature of brain process A may be defined as that which (a) one can discern, become aware of, only if A occurs in one's own brain (b) one can understand only if a brain process sufficiently similar to A has occurred in one's own brain. There is, I have argued, nothing conceptually puzzling about the nature of such a mental feature, and no *conceptual* problem as to how mental and physical features are inter-related. However, the feeling may persist that such problems must confront a brain process theory such as T_2 or Ti. I want now to suggest two main kinds of reasons why there may *appear* to be such conceptual problems even though in fact there are none.

In the first place, the assumption may be made that whenever a person experiences a sensation there is some entity which that person *perceives.* Thus it may be assumed that a person who is blindfolded and who experiences the visual sensation of a red square *perceives* some kind of red square. But (a) the person cannot consistently place or locate this red square among other perceived objects, (b) no one else is able to perceive the red square. Hence it seems that the red square is some kind of private, ghostly object, quite distinct from anything going on in the brain. It becomes wholly obscure how anything going on in the brain could possibly be related to this mysterious entity.

But of course a person who is blindfolded and who experiences the visual sensation of a red patch *perceives* nothing at all. Thus he has no evidence whatsoever for the existence of some entity with the *perceptual* properties of being red and square. It is true that he does have evidence for the existence of *something,* and is, furthermore, in a position to know

something about this 'thing': in view of the analogy with perception, we may, if we wish, talk here of a *kind* of observation, to be distinguished from perception, which might be called 'apperception'. We can say that a person X who experiences the visual sensation of a red square *apperceives* something with the *apperceptual* properties of being 'red_A' and square say. But 'What is happening to me is red_A' means roughly `What is happening to me is such that it is for me just as if I am perceiving a red object', where 'red' will be understood correctly by some other person Y only if whenever X and Y look at objects that X perceives to be red, sufficiently similar relevant brain processes occur in both brains. Hence there is no conceptual problem involved in ascribing such *apperceptual* properties to some kind of brain process. Further, it should be noted that in experiencing a visual sensation a person has evidence, not for the existence of some entity which can have no spatial location, but rather for the existence of some entity whose spatial location is left undetermined.

It is true that the mental feature of a particular brain process, A, can be detected, or discerned, only by that person in whose brain A occurs. But this fact only implies that there is something mysterious or inexplicable about the mental feature of A if one assumes in addition that the mental feature is some kind of perceptual or physical feature, which thus should be detectable or discernable by other people as well, by examining A under a microscope perhaps, or taking extremely delicate electroencephalograph recordings. But of course the mental feature of A is not any kind of perceptual or physical feature of A. A description of the mental feature of A tells one no more than what it is to have a physical process sufficiently similar to A occur in one's own brain: it tells one nothing about the perceptual or physical features of A.

The fact that apperception has a kind of certainty not possessed by perception may suggest that there is something conceptually puzzling about construing apperception as a *kind* of observation. Thus the possibility of verifying propositions apperceptually with certainty seems to contradict Einstein's eminently sensible dictum: In so far as a proposition refers to reality, it is not certain; in so far as it is certain, it does not refer to reality[9]. But if 'proposition' is construed as 'statement with a publicly agreed meaning' then no such contradiction arises. For even if there is one person who cannot be mistaken about the truth of, let us say, `I am experiencing the visual sensation of a red patch', there is no other person who can be *certain* as to what precisely this statement means, since in order to check on the meaning of 'red', elaborate comparisons of brain processes will have to be undertaken. I conclude that there is nothing

---

[9] Einstein's remark referred specifically to *mathematical* propositions. See **A.** Einstein, *Geometrie and Erfahrung,* Springier, Berlin, 1921, pp. 3-4.

conceptually puzzling about construing apperception as a kind of observation.

In the second place, it may be thought that there must be something conceptually very puzzling about the mental feature of a brain process if that feature cannot be described or understood in terms of any physical theory. Here is a property, so it seems, which must mysteriously evade the physicist's grasp, however strenuously he attempts to describe and understand it.

But of course the mental feature of a brain process evades being described by physics for the straightforward reason that it is just the kind of property that a physical description is specifically designed to avoid mentioning. Only if it is presupposed that a complete physical description must be a *complete* description will the fact that physics does not describe sensations *as sensations* seem conceptually disquieting.

Again, it is true in a sense that there are facts about brain processes which cannot be understood in terms of any physical theory. But this does not mean that these facts present the physicist with an insurmountable problem. It means rather that even if the physicist has solved all his problems, there might still be facts that he would not understand for the simple reason that he could not understand the *meaning* of any description of these facts.

According to both versions of the brain process theory developed here, T_2 and T_i, there is of course the genuine *empirical* (but non-physical) problem of correlating particular kinds of sensations with particular kinds of physical processes occurring in the brain. There is however no further *conceptual* problem concerning this correlation. Such a conceptual problem may appear to arise because it is assumed that if the correlation between a particular kind of brain process and sensation is to be understood it must be possible to give a *physical* explanation of this correlation. But such an explanation is impossible, not because this is a problem insoluble to physics, but because this is not a *physical* problem at all. In so far as it *is* a problem, it is the above empirical, non-physical problem; in so far as it is a *physical* problem, it does not exist. Hence once the above empirical problem has been solved, all that there *is* to understand about the corrrelation has been understood.

It is of course conceivable that knowledge of the physical structure of the brain alone should enable one to predict whether or not a person will be able to discriminate between the occurrence of two distinct kinds of physical processes, A and B, in his brain. This we may interpret as giving a physical explanation of why similar, or dissimilar, sensations correspond to A and to B. Such an explanation must however fail to render intelligible on what basis the person discriminates between the occurrence of A and B (if this is what he does do) : it must fail to make clear what sensations

the person experiences when A and B occur in his brain. (This can of course only be discovered if A and B occur in one's own brain.)

Finally, there may appear to be a conceptual problem concerning the relation between sensations and brain processes because on the one hand it seems clear that sensations are causally connected with certain physical processes (for example with physical processes occurring in the optic nerve), and yet it seems to be impossible in principle to give a physical specification of the causal laws involved. Physical processes can, it seems, only be causes of further physical processes; the nature of the causal link between the brain process and the sensation becomes wholly obscure.

As far as physics is concerned, one says that one state of affairs, A, is the cause of a second state, B, if and only if, from a description of A (initial conditions) plus a specification of relevant physical laws (physical theory) one could in principle deduce a description of B. We have every reason to believe that processes occurring in the optic nerve, for example, frequently are a cause of visual sensations, in this sense of 'cause'. However the description of a visual sensation that one could perhaps in principle derive from (a) a description of physical processes occurring in the optic nerve, (b) specification of other relevant initial conditions, (c) specification of relevant physical laws, would be a description of the visual sensation as a *physical process;* one could not, even in principle, derive a description of the visual sensation as a *visual sensation* from a purely *physical* specification of initial conditions and relevant physical laws. But this does *not* mean that there is some mysterious, additional causal law linking brain process with visual sensation, because of course the brain process *is* the visual sensation. In other words the occurrence of sensations can, in principle, be explained causally, i.e. physically. The fact that such an explanation invariably describes the sensation as a *physical process,* and not as a *sensation,* does *not* mean that there is some further conceptually puzzling causal link between physical process and sensation. In fact in order to derive a description of the sensation as a *sensation* from physical statements which describe relevant initial conditions and physical laws, one only requires, in addition to these physical statements, a statement of the empirical, *non-physical* fact that such and such a physical process *is* such and such a sensation.

University College, London.