

[Place, U.T. (2000) Identity theories. In M. Nani and M. Maraffa (Eds.), *A field guide to the philosophy of mind*. Rome: Roma Tre University. Retrieved Februari 9, 2019 from <http://host.uniroma3.it/progetti/kant/field/mbit.htm>]

Identity Theories

Ullin T. Place  - formerly University of Leeds

The Mind-Brain Identity Theory is the name usually, if somewhat misleadingly given to that form of psycho-physical materialism which holds, not that mind *qua* substance is the same independently existing substance as that anatomically distinguishable part of the human body known as the brain, but that mental events and mental processes are the very same events and processes as those patterns of brain activity with which they are known or strongly suspected to be correlated. It was only when it was stated in this form that psycho-physical materialism first became accepted as a defensible philosophical position. Nevertheless a quarter of a century was to elapse between the original statement of the mind-brain identity theory in the 1930's and its acceptance as a defensible philosophical position in the late 1950's.

I. Historical preliminaries

Boring

The earliest statement of the identity theory under the title was in Boring's book *The Physical Dimensions of Consciousness* published in 1933 . In that book Boring states his view as follows: "To the author a perfect correlation is identity. Two events that always occur together at the same time in the same place, without any temporal or spatial differentiation at all, are not two events but the same event. The mind-body correlations as formulated at present, do not admit of consideration as spatial correlation, so they reduce to matters of simple correlation in time. The need for identification is no less urgent in this case". (p.16). Boring's view did not command the serious attention either of philosophers or psychologists at the time for a number of reasons. As far as the psychologists were concerned behaviourism as a solution to and as a way out of methodological and philosophical problems within psychology had not been fully exploited and its limits and limitations discovered. As far as the philosophers were concerned, apart from the fact that Boring was writing as a psychologist for psychologists, the time was not yet ripe for the incorporation of such a revolutionary doctrine. The problem of identity and of referential identification had not yet moved into the forefront of discussion among the logicians. Frege's work on these topics

was virtually unknown outside the confines of the Vienna Circle. Phenomenalism in the form of the so-called *sense datum* theory was riding high and logical behaviourism had hardly been conceived, let alone explored as an alternative solution to the mind-body problem. Boring moreover, was himself apparently committed to combining the identity theory with a phenomenalist account of sensory qualities which on Leibniz's principle of the *Identity of Indiscernibles* would commit him to the view that certain brain events are literally green, high pitched, warm, sour or putrid, which for a philosopher would constitute an immediate knock-down *reductio ad absurdum* of his position.

The philosophical background

Twenty five years later the philosophical climate in the English speaking world was very different. The crucial influence was that of the later work of Wittgenstein as set out in his posthumous *Philosophical Investigations* published in 1953. In the United States parallel developments were taking place amongst those members of the original Vienna circle like Carnap and more in particular Feigl .

This development had a number of consequences which were crucial for the subsequent restatement of the Mind-Brain Identity theory in the late 1950's. In the first place it led to a revival of interest in the work of Frege and in his distinction between sense and reference. This distinction is fundamental to the doctrine of contingent identity as developed by Feigl in his (1958) and by Smart in his (1959). Frege's work on sense and reference also underlies the interest in the problem of how we identify the referent of a descriptive expression or proper name and thus to Strawson's view which draws attention to the importance of spatio-temporal location in this respect and which seriously undermines the traditional view of the mind as an independently existing spiritual substance. This interest in the problem of referential identification in the context of the use of language for the purposes of inter-personal communication is the basis of Wittgenstein's *private language argument*. This argument of Wittgenstein's together with Austin's critique of the argument from illusion in his *Sense and Sensibilia* led to the dramatic collapse of phenomenism and the Berkeleyan form of Idealism which is associated with it as defensible philosophical positions within the British philosophical tradition in the period immediately following the Second World War. The refutation of phenomenism and the recognition that naive realism, the view that what we directly perceive are real spatially extended material objects in a three dimensionally extended material world, is a viable philosophical position was an essential prerequisite for the development of a philosophically defensible form of the mind-brain identity theory, since it is only when we recognise that the language we use to describe our private experiences and sensations is a metaphorical extension of a language whose basic function is to describe material objects and their properties as they exist and occur in a three dimensionally extended spatial world, that we can circumvent the objection that experiences have properties such as greenness, high pitch, warmth, sourness and putridity that no brain process could conceivably

have. We have already suggested that it was this adherence to phenomenalism that led to the failure of Boring's original statement of the identity theory to gain widespread acceptance. It was also we would suggest, his failure to break away effectively from his deep-rooted phenomenalist prejudices which led to Feigl's recantation of his 1958 statement of the identity theory in the introduction to the reissue of that paper as a separate volume in 1967.

Wittgenstein's private language argument was also important as a factor in the development of a philosophically viable form of psycho-physical materialism in so far as it demonstrates that any language which is capable of serving as a medium for inter-personal communication must necessarily presuppose the existence of a spatially extended material world to which its basic concepts are referentially anchored. This leads to a rejection of the egocentric epistemology of Descartes and thereby undermines the formidable Cartesian argument for an independently existing spiritual substance. It also helped philosophers to recognise for the first time in three hundred years that the primary function of the psychological concepts of ordinary language is not to enable the individual to describe his own private experiences, but to enable him to characterise the mental capacities and propensities of other people and to explain the behaviour of others in terms of these capacities and propensities. Once we begin to ask what it means *for him* to know, believe, want or intend something instead of asking what it means *for me* to know, believe, want or intend something, *logical behaviourism* at least as an account of these mental state concepts, becomes almost irresistible. Wittgenstein himself gave a successful logical behaviourist analysis of what it means to understand something in terms of the ability to 'go on' correctly (1953, I, §143-155), as well as his less successful attempt at a logical behaviourist account of sensation (1953, I, §244).

However more important than his own specific contributions to the logical behaviourist view, was Wittgenstein's undoubted influence inspiring Ryle's exposition of this point of view in *The Concept of Mind*. Strictly speaking logical behaviourism, if taken to its logical conclusions, is an alternative and rival to materialism as an account of the mind-body relationship. For if and in so far as all we are doing when we use mental concepts is talking in a logically complex way about the publicly observable behaviour of human beings, it follows (a) that there is no separate class of mental states and events over and above behavioural events and behavioural dispositions and (b) that the relationship of mind to brain is simply a matter of the relatively unproblematic causal relationship between brain activity and overt behaviour. Nevertheless had it not been for the fact that Place and Smart had both been strongly influenced by Ryle's logical behaviourism and had both been impressed with how close this view comes to getting rid once and for all, of the private world of mental events, neither Smart (1959) nor Place (1956) would have ventured to explore the possibility that those aspects of mental life which had proved impermeable to the dispositional analysis which Ryle had used with such devastating effect elsewhere, might perhaps be rescued from the strange extra physical limbo to which Descartes had

consigned them by postulating their identity with events and processes in the brain.

The Restatement of the Mind brain identity theory

The mind-brain identity theory in the form in which it became accepted as a serious philosophical thesis stems from three papers published in the late 1950's. The earliest of these was Place's paper 'Is consciousness a brain process?' (1956). This was followed in 1958 by Herbert Feigl's paper 'The "Mental" and the "Physical"' and in 1959 by J.J.C. Smart's paper 'Sensations and Brain Processes'. Although there are certain differences of detail in the positions adopted in these three papers, the area of agreement was sufficiently great for all three of the original protagonists to be able to agree that they were all defending the same basic position. In the case of Place and Smart there was a direct personal connection in that Place had developed his view on the basis of a series of discussions in which Smart had participated which took place in Smart's Department in the University of Adelaide, of which Place was then a member in 1954. He had already announced his intention to defend the thesis that "the logical objections to the statement 'consciousness is a process in the brain' are no greater than the logical objections which might be raised to the statement 'lightening is a motion of electric changes'" (p.255) in his paper 'The concept of heed' which appeared in the same year. But it was only after this series of discussions with Smart, C. B. Martin and D. A. T. Gasking (while on a visit to Adelaide from the University of Melbourne) that the argument of 'Is consciousness a brain process?' was finally knocked into shape. During these discussions, Smart although he became increasingly sympathetic to the Place point of view as time went on, had not entirely accepted the position which we later came to call 'the identity theory'. On Fall term 1957 however, he was giving a graduate class at Princeton on Ryle and Wittgenstein during which he came round to agreeing with Place, advocating the Identity view and inviting objections to it. His 1959 paper records these objections and develops his answers to them.

Feigl

Feigl's version of the identity theory was developed quite independently of Place and Smart and grew out of his (1950). He argued in that paper "that the designata of the mentalistic language are identical with the descripta of the behaviouristic language and that both are identical with the designata of the neuro-physiological language". In his 1958 paper Feigl allowed the "descripta of the behaviouristic language" to drop into the background and attempted to specify more precisely those concepts within the mentalistic language whose designata could reasonably be supposed to be identical with certain events or processes in the brain. "The word 'mental' in present day psychology", he argued "covers ... not only the events and processes of direct experience (i.e. the raw feels) but also the unconscious events and processes, as well as the 'intentional acts' of perception, introspective awareness, expectation, thought, belief, doubt, desire, volition, resolution etc. ... since *intentionality* as such is to be analysed (on Feigl's view) ... in

terms of pure semantics (and thus falls under the category of the *logical* rather than the psychological), it would be a category mistake of the most glaring sort to attempt a neurophysiological identification of this aspect of mind." (p.445). In the light of these considerations he concludes: "the identity thesis which I wish to clarify and to defend asserts that the states of direct experience which conscious human beings 'live through' and those which we confidently ascribe to some of the higher animals, are identical with certain (presumably configurational) aspects of the neural processes in those organisms" (p.446).

II. The Identity relation

Similarities and Differences between the positions of Place, Smart and Feigl

Although Place, Feigl and Smart all agreed that they were defending the same position for which they all sooner or later came to describe as the mind-brain identity theory, there are inevitably certain detailed differences between the different presentations of what is basically the same thesis. But before discussing these differences it will be helpful to set out the points on which from the outset, there was agreement between all of them. They agreed on the following four points:

1. The private experiences or sensations of the individual are reducible without remainder (are nothing but) certain as yet unspecified events or processes in the brain (the identity thesis).
2. The identity thesis is a contingent proposition i.e. it is not a logically necessary truth. Descriptions of an individual's private experience do not have the same sense or meaning as the physiologist's description of the hypothetical brain processes in which the private experiences themselves in fact consist.
3. The truth of the identity thesis is at least partly a matter of empirical determination.
4. The identity thesis applies only to certain aspects of mental life - consciousness (Place), the raw feels of experience (Feigl), sensations (Smart). The cognitive and volitional (intentional) aspects of mental life are *not* reducible to brain states or processes; but are (conceptually) reducible to some kind of semantic, logical or verbal competence or propensity.

Identity versus equality

The differences between three formulations of the same basic thesis can all be related to the above mentioned four points of agreement. Despite Place's description of the symmetrical identity relation in terms of the asymmetrical relation of composition to be discussed in a moment, all three Identity theorists share the same 'model' of identity. It is important to

contrast two concepts of identity which we may call the Leibnizian and Fregean models respectively. On the Leibnizian model, the paradigm case of an identity statement is a tautology of the form $x=x$. An identity statement is a statement which is either of this form or reduces to this form by the process of mathematical substitution. Such statements are necessary truths. It makes no sense on this model to speak of 'contingent identity'. Leibniz, of course, famously thought that all true propositions, whether necessary or contingent are ultimately reducible to identity statements in this sense. But this was because he believed that each individual entity has a concept containing all the predicates that are timeless true of that individual throughout its career through time. Not only does this view entail a metaphysics which most of us find difficult to accept, it undermines the distinction between the necessary and the contingent which Leibniz himself had done so much to clarify. What we are calling the Fregean model of identity (though without making any claim as to whether Frege himself would have subscribed to it in exactly this form), and which is at work in the brain-mind identity formulations, takes as its paradigm case of an identity statement a statement such as 'Bill Lycan is the American philosopher Ullin Place met for the first time in Sydney in 1983'. On this Fregean model, identity is construed as a relation between two different names or descriptions such that they both refer to the same individual. In contrast to identity statements of the Leibnizian variety such statements are typically contingent both in the sense that it is not contradictory to suppose them to be false and in the sense that they just happen to be true. Note that a proposition of the form $x=x$ does not even qualify as an identity statement in the Fregean sense. This is (a) because the application of the '=' sign in such cases does not depend on the 'reference' in the Fregean sense (Bedeutung) of the signs it connects and (b) because the signs that are so connected are two tokens of the same sign not two different signs. Suppose I have four and only four coins in my left hand trouser pocket and four and only four coins in my right hand trouser pockets. In this case the number of coins in both pockets is the same. The formula $a=a$ is satisfied. But that equality is not identity. It is numerical equality. <1>

Identity versus composition

In relation to the statement of the identity thesis itself (Point 1) there is a difference between Place's version of the thesis which was stated in terms of what he referred to as "the 'is' of composition" (1956) and the version of Feigl (1958) and Smart (1959) who followed Boring (1933) in speaking of the identity of private experiences or sensations, on the one hand with brain processes or brain events on the other. Place's contention was not that consciousness is identical with or the very same thing as the brain processes with which it is correlated, but that consciousness consists entirely in or is entirely composed of brain processes. In other words he construed the relationship of experiences to brain processes in terms of the substantial micro-reduction of a substance into its constituent parts at a lower or more microscopic level of analysis, whereas Feigl and Smart

discussed the relationship in terms of Frege's notion of the identity of the referent of two descriptions.

Place now admits that for a long time he was in a state of some confusion as to whether or not this difference was merely a matter of two different ways of saying the same thing or whether there was not perhaps, as he once argued (1967) a reason for preferring his formulation in terms of composition to the more generally accepted formulation in terms of identity. In favour of the view that two formulations are two ways of saying the same thing is (1) the fact that both are equally effective in eliminating mental events and mental processes as a separate class of events and processes with no extension or location in ordinary three dimensional space and (2) the fact that both formulations imply that the relationship is a contingent one. Substantial, material or micro-analysis and reduction is a matter of breaking down the entity itself into its constituent parts in contrast to conceptual analysis and reduction which is a matter of breaking down the concept under which an entity falls into its component conceptual elements. Such a micro-reductive analysis presupposes a closer more rigorous, more detailed and more scientific investigation of the entity or stuff which is being analysed, than is required in order to identify it as an entity or stuff of a certain kind at the macroscopic level. Empirical observation is likewise required in order to demonstrate that two logically independent descriptions, as in Frege's example of 'the Morning Star' and 'the Evening Star', have a common referent (the planet Venus) which is employed both by Feigl (1958) and by Smart (1959) in their expositions of this view.

The principal argument against treating the two formulations as equivalent is the argument that the relation of identity is *symmetrical* in that if A is the same thing as B, it follows necessarily that B is the same thing as A. The composition relation on the other hand, is *asymmetrical* in that if A is entirely composed of B's, we cannot say that B's are entirely composed of A. What we have to say is that B's entirely comprise or make up A. Furthermore, in the case of material analysis and reduction, while it is true that an entity can be said to be entirely made up of the parts of which it is composed, a mere collection of all the parts of a thing does not comprise the thing of which they once formed part, unless they are arranged in the particular way in which they were originally arranged. A list of parts however complete, is only a description of the same thing as is described by a description of the whole of which they form part in so far as their form and arrangement is also specified. Where both form and matter are specified in the description of the analysis, there is no doubt a sense in which the two descriptions refer to one and the same thing; but the relationship is still in an important sense, *asymmetrical* in that the description of the analysis of thing into its constituent parts tells us a great deal about it which the macroscopic description does not mention and the microreductive description *explains* the characteristics of the macroscopic entity in a way that the macroscopic description cannot be said to explain the micro-description.

Nevertheless, although there is still an element of asymmetry between the macroscopic and the microscopic description whereby the microscopic explains the macroscopic and not vice versa, this is not the sort of asymmetry which is incompatible with asserting the symmetrical relationship of identity as far as the common referent of the two descriptions is concerned. Provided we specify their form and arrangement we can equally well say that the parts of a thing so arranged are the same thing as the thing itself and that the thing itself is the same thing as the collection of its parts is arranged in that particular way.

But although this argument shows that there is no logical incompatibility between these two formulations of the relationship, it is clear that the two formulations are different and complementary to one another. The formulation in terms of composition has the advantage of drawing our attention to examples such as the cloud's consisting of water droplets, the lightning's consisting of an electric discharge through the atmosphere or water's consisting of H₂O, where we identify an entity, a process or a stuff with its scientific microdescription which provide a much closer parallel to the experience-brain process case than does the standard example of the contingent identity of the referent of two logically independent descriptions - the Morning Star/Evening Star case. On the other hand the fact that a collection of the parts of which a thing is composed is not by itself the same thing as the original entity, process or stuff might be thought to allow too much latitude for the supernatural mind-stuff to creep back in through the back door under the banner of the emergent whole being greater than the mere sum of its parts. This back door the identity formulation keeps firmly closed.

It is worth pointing out in this connection that the reason why the standard example of two logically independent descriptions having a common referent are not closely comparable to the experience/brain process case is that they all involve different descriptions which are true of and can be used as a means of identifying a single particular individual. This is true both of Leibniz's case where the same individual is identified either by the proper name 'G. Julius Caesar' or by such descriptions as 'the man who crossed the Rubicon' or 'the man who was murdered by Brutus and his associates', of Frege's case where, the same individual heavenly body is identified either by the proper name 'Venus' or alternatively by the descriptions 'the Morning Star' and 'the Evening Star' and of Place's case (1956) of an entity which is describable both as a table and as an old packing case. In the experience/brain process case, by contrast what is asserted is the identity not of one particular common referent of two or more descriptions but of the referents of two whole classes of descriptions, descriptions of private experiences in general on the one hand and the neurophysiological descriptions of all the corresponding brain processes on the other.

Token versus Type Identity

Both *This table is an old packing case* and *The Morning Star is the same object as the Evening Star* are cases of token identity, cases where two

descriptions with different senses just happen to apply to one and the same particular object. Such cases are extremely common. Indeed any non-analytic proposition which asserts the co-application of two conceptually unconnected predicates of the same object is of this kind. But so are all those which apply in the case of an aggregated collection of objects, such as the coins in my pocket which all happen to be copper. But the target case in Place's paper, the claim that consciousness is a process in the brain, is not like this. Hence we have two types of thing, consciousness and a certain as yet unspecified type of brain activity, which don't just happen to satisfy two descriptions, but which are such that the features which lead us to apply the one description also leads us to apply the other, and where the absence of the same features would in all cases lead us to withdraw both. This, in other words, is a typical case of type- rather than token-identity. But whereas the typical token-identity statement 'His table is an old packing case', if true, is contingent and synthetic, the typical type-identity statement of which 'Water is H₂O' is a paradigm case is necessary and analytic (for further elaboration see Place 1991).

Compositional type-identity statements as necessary and analytic

In his 1997 revision of the (1956) paper, Place points out that "if we lived in a world in which all tables without exception were packing cases, the concepts of 'table' and 'packing case' in our language would not have their present logically independent status. In such a world a table would be a species of packing case in much the same way that red is a species of colour. It seems to be a rule of language that whenever a given variety of object or state of affairs has two characteristics or set of characteristics, one of which is unique to the variety of object or state of affairs in question, the expression used to refer to the characteristic or set of characteristics which defines the variety of object or state of affairs in question will always entail the expression used to refer to the other characteristic or set of characteristics." (1997, p. 7). In other words in a universe where it is both true and obviously true that $(x) (Fx \text{ if and only if } Gx)$, $(x) (Fx \iff Gx)$ will become true by definition such that anything that does not have the characteristic F will not be accepted as a case of a G. One consequence of this principle is that we can only expect to encounter cases where a class of things has two properties whose descriptions are logically independent of one another when the fact that both descriptions apply to the same class of things is not apparent at the level of common sense knowledge and observation. It would seem moreover, that the only cases where this happens are ones where previously unknown properties of familiar things are revealed by scientific micro-analysis. Another consequence of this principle is that when, as in cases like water and H₂O, the substantial analysis of a class of entities or stuffs becomes a matter of common knowledge the statement that water is a compound of two atoms of hydrogen to one of oxygen ceases to be a contingent proposition and becomes true by definition, so that anything that does not have this chemical composition is no longer accepted as a genuine case of water. As Place has argued in his (1967, footnote pp.66-7), we may expect a similar conceptual development to take place in the case of the mind-brain

relationship once the identity of experience and brain processes becomes a matter of known and accepted scientific fact.

III. Objections and replies

Mind-brain identity and empirical evidence

In his (1956) Place argued that the thesis that consciousness is a process in the brain is "a reasonable scientific hypothesis, not to be dismissed on logical grounds alone". (p. 44). If the proposition 'experiences are processes in the brain' is a contingent proposition and not a necessary truth and if moreover, it is a true contingent proposition, since a true contingent proposition is one which can be denied without self-contradiction or without contradicting certain self-evident premises from which it is deduced, it would seem to follow that the only grounds we can have for asserting it to be true must be the evidence of scientific observation. However since the empirical evidence in favour of such a view though impressive, is not yet by any means conclusive, we cannot yet regard the identity thesis as a matter of established scientific fact. Hence the description of it as a scientific hypothesis.

Smart (1959) argued that this contention "is partly right and partly wrong. If the issue is between (say) a brain-process and a heart thesis, or a liver thesis, or a kidney thesis. The right sort of things don't go on in the heart, liver or kidney nor do these organs possess the right sort of complexity of structure. On the other hand, if the issue is between a brain-or-heart-or-liver-or-kidney thesis (that is, some form of materialism) on the one hand and epiphenomenalism on the other hand, then the issue is not an empirical one. For there is no conceivable experiment which could decide between materialism and epiphenomenalism." (p.155). Smart then goes on to suggest that the only way to settle the issue as between materialism and epiphenomenalism is by appealing, as Boring (1933) had done before him, to "the principles of parsimony and simplicity", in other words, to Ockham's razor, by which principle needless to say, materialism wins hands down, since it not only reduces the number of separate entities which have to be postulated in order to account for the mind-brain relationship by half, it also removes the necessity of having to postulate a unique set of entities existing outside the three dimensionally extended spatial universe of science and commonsense, standing in a highly problematical causal relationship with events in that universe.

Place in his rejoinder (1960) to Smart's paper argued that there are certain logical conditions which must be satisfied to enable us to say that a process or event observed in one way is the same process or event as that observed in (or inferred from) another set of observations made under quite different conditions:

For the purposes of the present argument it does not matter whether this account of the logical criteria used to establish the identity of an event described in terms of two procedures of observation is correct or not. What is important is that there

must be some logical criteria which we use in deciding whether two sets of correlated observations refer to the same event or to two separate but causally related events. The problem of deciding what these criteria are is a logical problem which cannot be decided by experiment in any ordinary sense of the term; and since we cannot be certain that the criteria are satisfied in the case of sensations and brain processes unless we know what the criteria are, the issue is to that extent a philosophical issue. Moreover, even if we agree on the nature of these logical criteria, it is still open to the philosopher to question the logical propriety of applying them in the case of sensations and brain processes.

For the sake of argument however, let us assume that these philosophical issues have been settled and that they have been settled in favour of the materialist hypothesis. We now find ourselves faced with a purely empirical issue namely, whether there is in fact a physiological process, be it in the brain, the heart, the liver, the kidney or the big toe, which satisfies the logical criteria required to establish its identity with the sensation process. As it happens we already know quite enough to be quite sure that if there is such a process, it must be situated in the brain and even within the brain there are extensive areas that can be ruled out with virtual certainty as possible loci of consciousness - areas for example, where brain lesions produce motor disturbances without any change in consciousness other than an awareness of the disability itself and emotional reactions to the problems it creates. But the empirical problem is not, as Smart seems to think, simply a matter of determining the precise anatomical location of this physiological process. Even assuming that we know what these criteria are and are satisfied that they are applicable in this case, we cannot regard the question as finally settled until a process satisfying the necessary criteria has been discovered or until we are sure that we know enough about the brain to be certain that no such process exists. (pp.102-3)

An additional point to add to this part of the argument is that just as there must be logical criteria for deciding whether we are dealing with two sets of observations of one and the same process or event or with two sets of observations of two separate but correlated events, so also there must be logical criteria for deciding in the second case whether the two independent but correlated processes and events are causally connected and if so, in which direction the causal relationship operates. Given that we can agree as to what these criteria are or should be, it would seem to follow that empirical evidence is relevant not only in deciding the issue as between materialism and dualism, but also assuming that the materialist solution can be excluded in deciding the issue as between interactionism, epiphenomenalism and psycho-physical parallelism.

The logical criteria of identity

In the light of these considerations it is evident that the crucial issue that must be decided before we can proceed to look for a process in the brain which must be plausibly identified with the conscious experiences described and reported by the human subject, is the nature of the logical criteria which we use and are justified in using in deciding whether two independent sets of observations are observations of two different sets of correlated events or of one and the same process or event. This is an issue which can be approached in two different ways. One approach, which is one usually favoured by logicians and philosophers is to argue *a priori* from the nature of the notion of identity itself, as defined in terms of Leibniz's Principle of

the *Identity of Indiscernibles*. The other approach is to argue empirically or inductively from examples of other cases where we accept two independent sets of observations as observations of the same event or process, as compared with cases where we treat the two sets of observations as implying the existence of two separate, but correlated events or processes.

A priori argument from Leibniz's Principle

Leibniz's principle of the Identity of Indiscernibles holds that if all the predicates that are true of an entity A are also true of what is taken to be another entity B and all the predicates that are true of B are also true of A, then A and B are not two things but one and the same thing. As applied to the case of experiences and brain processes what this principle means is that if we were to discover a brain process which had all the properties which the introspecting subject attributes to his current experience and had no properties which the introspecting subject was not prepared to acknowledge as being true of his current experience, we would be forced to conclude that the experience and the brain process are one and the same thing. In practice of course, no such brain process has yet been discovered. Furthermore, we already know enough about brain processes to be quite certain that no such brain process ever will be discovered. For although it is both conceivable that we shall eventually discover a brain process which has all the properties which the introspecting subject attributes to his experience, there are a number of properties which apply to all brain processes which the introspecting subject would never think of predicating of his experiences solely on the basis of his having or experiencing them. One such property is the property of involving the firing of at least one and probably many thousands of neurons each of which has a specific location within the anatomical structure of the brain.

Now the fact that any brain process which is capable of being studied by the neurophysiologist necessarily has a number of properties which the introspecting subject would never think of predicating of his experiences is not enough to show that the experience and the brain process are two different things and not one and the same thing. For there is no reason why an experience should not be supposed to have properties which its owner does not know that it has. What it does mean however, is that we can never hope to use the indiscernibility of what is observed under the two sets of conditions as an argument for their being one and the same thing. That is why Leibniz's principle of the Identity of Indiscernibles has been appealed to far more often in arguments *against* the mind-brain identity theory than in arguments in its favour.

Objections to the identity theory from Leibniz's Law

In using Leibniz's Principle of the Identity of Indiscernibles in formulating objections to the mind-brain identity theory an appeal is made to the converse of Leibniz's principle, the principle usually referred to as *Leibniz's Law*. Leibniz's Law, as distinct from the Principle of the Identity of Indiscernibles from which it derives, holds that if two descriptions A and B

refer to one and the same entity, then any predicate which forms a true proposition when predicated of A must also form a true proposition when predicated of B. In order to generate an objection to the mind-brain identity theory based on Leibniz's Law in this sense, what has to be shown is, either that there are predicates which form true propositions when predicated of experiences which form false propositions when predicated of brain processes, or that there are predicates which form true propositions when predicated of the brain processes whose occurrence is correlated with reports of experiences which form false propositions when predicated of the experiences themselves.

Now since the relevant brain processes have not yet been identified, even tentatively, we are not in a position to use the empirical evidence to show that the experiences have properties which the brain processes do not have and *vice versa*. Those who have put forward this kind of objection to the identity theory have therefore, looked not for predicates which are true of the one which are as a *matter of fact* false of the other, but for predicates which are true of the one which are *necessarily* false when applied to the other by virtue of some logical contradiction which allegedly arises when a predicate that is true of an experience is predicated of a brain process or when a predicate that is true of brain processes is predicated of an experience. It is not sufficient to show merely that these experiences have properties which we would not think of predicated of brain processes if we did not suspect them of being one and the same thing, or that brain processes have properties which we would never think of predicating of experiences, if we did not suppose them to be one and the same. For the fact that something is not known to have a given property is no evidence that it does not have that property. It might of course, be argued that it is a peculiar and distinctive feature of experiences that they can only have these properties which their owner knows that they have; but it is only necessary to think of the case of the pain which has the property of being caused by some internal lesion whose nature and existence is entirely unknown to the sufferer to refute that suggestion.

Nor is it enough to show that it sounds odd or unnatural to say, for example of an experience that it consists of the firing of several million neurons in the cerebral cortex or of a brain process that it has a roundish reddish look. It is true that we often use the oddness or unnaturalness of combining a given subject term with a given predicate as *prima facie* evidence of a logical contradiction between the two terms. Such evidence however, is far from conclusive. All it shows is that the sentence in question is not one that we normally have occasion to utter. This may be because of a logical contradiction between the subject and predicate terms; but it may also be that the apparent incongruity is due to nothing more than our unfamiliarity with the notion that two things which have long been assumed to be different are in fact one and the same. If this is the only reason and if the identification becomes accepted as a useful scientific assumption, we may confidently predict that such locutions as those which attribute to experiences those predicates which belong primarily to the brain processes with which they have been identified will come to seem very much less

incongruous in the future than they do now. At one time, it must have appeared very odd and incongruous in terms of our ordinary ways of talking to speak of light as consisting of waves and of differences of colour in terms of the differences in wave length of the light emitted from or reflected by the object in question, though now such conceptions are so familiar to us that they require no justification.

This explanation of the incongruity of some of the things we should be compelled by Leibniz's Law to say, if the identity hypothesis is accepted, can only be applied however, in the case of those predicates of the brain processes which are applied to experiences by virtue of the acceptance of the identity hypothesis. We cannot explain and resolve incongruities which arise when we apply the predicates of experience to the brain processes in this way. The reason for this is related to the asymmetry in the identity relation between the two terms in such cases which we discussed above in relation to the formulation of the relationship in terms of micro-reductive composition rather than identity. In these micro-reductive scientific identifications the relationship is asymmetrical even in the case where the form as well as the matter is included in the micro-reductive account, in that the scientific micro-reduction explains the description that is given at the pre-scientific macroscopic level and not *vice versa*. This means that, whereas the micro-description, if it is to provide as it must do, a complete explanation of the macro-description, must mention all the properties mentioned in the macro-description, it must also if it is to provide a genuine explanation, mention properties of the entity or process in question which are not mentioned in macro-description. Consequently there will be properties attributed to the common referent under its macro-description by virtue of its newly discovered micro-description whose predication of the macro-description will initially appear incongruous solely by virtue of unfamiliarity. On the other hand, any incongruity between the micro-description and the predicates applied it by virtue of their application to the common referent under its macro-description must *prima facie* be interpreted as a failure of the micro-description to provide an adequate account of the entity to which the macro-description refers. We could only conclude from such a failure that the micro-reductive account, so far given is incorrect and hence, in the case of some property of experience that remains obstinately incongruous when predicated of any currently conceivable variety of brain process that such an experience cannot be any kind of currently conceivable brain process.

However, when we examine those cases where there is an alleged logical incongruity or contradiction involved in predicating of a brain process some property of an experience, it turns out either that the property in question is a phenomenal property whose 'topic neutral' character has not been appreciated (see below), or that it is an intentional property which applies to things only under one description and not under any and every description which is true of the entity in question and whose application to the common referent of two descriptions under both of them is not therefore required by Leibniz's Law.

The principles outlined above can be illustrated by reference to three cases of alleged infringements of Leibniz's Law to which the critics of the Mind-Brain Identity Theory have drawn special attention. These are: (1) the phenomenal properties of experience, particularly their colour properties, (2) the necessarily private character of conscious experiences and (3) the lack of any clear cut spatial location in the case of private experiences.

Phenomenal Properties

The argument from the phenomenal properties of experience may be stated as follows: experiences have phenomenal properties such as the property of being green, red, blue or yellow. It makes no sense to describe a brain process as green, red, blue or yellow. Hence experiences have properties (phenomenal properties like being of a certain colour) which no brain process can have. Hence by Leibniz's Law experiences cannot be the same thing as brain processes.

This argument breaks down because it ignores what Smart (1959) has called the 'topic neutral' character of the descriptions that we give our own private experiences. The term 'topic neutrality' describes a feature of introspective reports i.e.: the descriptions which the individual gives of his own private experiences, whereby such statements do not and necessarily cannot mention any actual properties of the experiences themselves. An experience on this view can only be described in terms of its resemblance to other experiences identified in terms of their standard publicly observable concomitants. According to Smart the underlying logical form of these phenomenal descriptions of experience is a sentence frame of the form 'There is something going on in me which is like what typically goes on in me when I look at, listen to, savour or feel something that is O ' or alternatively 'There is something going on in me which is like what typically goes on in me when I am inclined or tempted to O '.

If and in so far as our descriptions of our own experience are topic neutral in this sense, it follows that when we describe an experience such as an after image as green, we are not predicating the property of greenness to the experience itself, we are saying only that the experience is the sort of experience we normally have when we look at objects or perceptible phenomena which *do* have the property of being green. In other words visual experiences do not have the property of being literally green any more than brain processes have this property. Leibniz's Law is not infringed.

It is true of course, that visual experiences do have a property which you may call if you wish 'phenomenal greenness' or the 'phenomenal property of greenness', which is the property of the experience which we describe by saying either that there is something in the external visual environment that looks green or that it looks as if there were something literally green in the external visual environment. But it is not this kind of phenomenal greenness that we cannot predicate of brain processes. What brain processes cannot be is literally green, green in the sense that grass or traffic lights are green.

If however, all we mean when we say that some visual experiences are phenomenally green is that they are the kind of experiences that normally occur when we are visually presented with objects and light sources which are literally green and which enable us to recognise literal greenness as a property of such objects and light sources under normal conditions of illumination, then to say that an experience is phenomenally green is to say nothing about it which could not equally well be said of those brain processes in the visual cortex which enable us to discriminate between different coloured objects in the field of view.

This argument presupposes of course, that colour words like 'red', 'green', 'blue' and 'yellow' when used literally, refer to physical properties of objects and phenomena in the external physical world and are not, as has been traditionally supposed, the names of certain properties of the individual's subjective visual experience. That this presupposition is correct as an account of how we use colour in ordinary language is clearly shown by the fact that it is objects like leaves and blades of grass that we describe as green, not the experience we have when we look at such objects, by the fact that grass ceases to be green only when it withers and dies, not when the level of illumination is too low for its greenness to be seen, and by the fact that we draw a distinction between the colour that an object actually or really is and the colour it appears to be under abnormal conditions of illumination.

Despite these obvious linguistic facts the belief that colour properties are what Galileo and Locke called *secondary qualities* which, like 'beauty', are 'in the eye of the beholder' rather than in or on the surface of the object itself is very deep rooted not only amongst philosophers but also amongst physicists and physiologists who are concerned with the phenomena of light and colour discrimination. This prejudice has led Smart to devise a version of the topic neutral formula which will eliminate all reference to colour properties from descriptions of experience so that, instead of saying of a yellow after image that it is the sort of experience we normally have when looking at a patch of paint or patch of light that is really and literally yellow, we have to say "what goes on in me is like what goes in me when a lemon is in front of my eyes" (p.91). As M.C. Bradley has pointed out (1963), this use of the topic neutral formula to eliminate all reference to colour properties leads to a vicious circle. In Place's view it also brings the thesis of the topic neutrality of sensation reports into disrepute by presenting it, not as a plausible analysis of reports of experience as they occur in ordinary language, but as a sophisticated device for eliminating properties which cannot readily be fitted into a physicalist universe.

This attempt to eliminate colour concepts from our vocabulary is wholly unnecessary and indeed, undesirable from the standpoint of the mind-brain identity theory. Colours are straight forward physical properties of the exposed opaque surfaces of material objects, of the whole body of objects that are translucent or which are themselves light sources. Colour concepts are conceptually irreducible in that we cannot further define what it means to say of something that it is red, green, blue or yellow. We can only explain

what it means by pointing to examples. However in the light of modern physical theory, we now know that colour properties consist in (i.e. are reducible by substantial analysis to) the property of selectively reflecting, transmitting or emitting light radiation within a particular band of wave lengths. It is true that the particular range of wave lengths which we use in discriminating objects of different colours is determined, somewhat arbitrarily from a physical point of view, by the sensitivity of human retina. It is also true that the classification of the visible spectrum in terms of our various colour words is determined by the physiology of the colour reception processes in the retina, rather than by any objective physical boundaries along the continuum of light wave lengths within the spectrum. The rather fuzzy boundary between red and orange which we see when we look at an actual band of spectrally diffracted light is where it is and is experienced as a boundary of this kind only by virtue of the physiology of our visual discriminatory apparatus. But this does not mean that there is no actual boundary between red and orange light at this point. Given that red light is (in a contingent identity or composition sense of 'is') light of a wave length of between 0.72u and 0.63u and that orange light is similarly light of between 0.63u and 0.59u, there both necessarily and actually *is* a boundary between these two ranges of light wave lengths at a wave length of about 0.63u which is the wave length of the spectrally diffracted light at the point where it changes from red to orange. If the physiology of the normal human eye were different, we would not observe a boundary at this point, nor would we have the particular set of colour concepts that we do have. But if by some miracle the physiology of every normal human retina were suddenly changed so that everything which used to look red now looked green, it would not suddenly become true that everything that used to be red is now green. In terms of our present set of colour concepts, it is true, has always been true and always will be true, whatever happens in the future to the human eye, that the leaves of all currently existing varieties of living grass are green. If everyone suddenly became red-green colour blind, no one would any longer be able to distinguish between these two forms of light or between otherwise identical objects differing only with respect to the light of these two ranges of wave lengths which they selectively reflect without using complex instruments to measure the light wave lengths involved. Under these conditions the distinction between red and green would, no doubt drop out of our ordinary colour vocabulary; but in terms of our present vocabulary, grass would still be green and blood would still be red. The fact that there would no longer be any point in saying this would not make it any the less a true proposition.

The Privacy of Private Experience

The argument from the necessary privacy of private experiences was put forward as an objection to the mind-brain identity theory in a paper by Kurt Baier (1962). The argument may be stated briefly as follows. It is a defining property of private experiences that they are private in the sense that the owner of such an experience has what Baier calls 'Final epistemological authority' with respect to the nature and occurrence of his own experiences. He may make a mistake in his description of his own experience, but only

he can correct such a mistake. The owner of a private experience has a way of coming to know what his experiences are which is "not open to others". The nature and occurrence of his brain processes, on the other hand, while not a matter of straightforward public inspection, can be ascertained if anything, more readily by another person than by their owner. Thus private experiences have a property, namely that of being known privately, which brain processes do not have, while brain processes have a property, namely that of being publicly ascertainable which does not apply in the case of private experiences. Hence by Leibniz's Law, the two cannot be one and the same thing.

Smart's reply to Baier (1962) which focuses on the problem of the alleged incorrigibility of sensation reports and on the issue of what we would have to say in a case where the individual honestly reports the occurrence of an experience for which no counterpart in the brain can be discovered, does not meet the objection when stated as, we have stated it, as an infringement of Leibniz's Law. In his (1967) Place tried to argue that the objection can be met by treating the privacy of experiences as the lack of a property - public observability or ascertainability - which belongs to brain processes, rather than as a positive property of experiences. This would then enable us to account for the presence of this property in the case of brain processes and its absence in the case of experiences in terms of the principle of the asymmetry of the composition relation which allows the microdescription of an entity to have properties which would not be ascribed to it on the basis of the macro-description but not *vice versa*. However this resolution of the problem no longer appears to Place satisfactory, since it is clear that the special kind of introspective knowledge that we have of our own experiences is a positive property of them and not merely a matter of their not being available to public inspection.

The solution that Place now favours is based on the recognition that privacy is a matter of how we come to know about our experiences and that 'knowing' is an 'intentional predicate'. It is the defining characteristic of an intentional object that it has a description or proper name for which we cannot legitimately substitute another name or description by virtue of an identity statement. If Joe knows that James has red hair and James is the brother of John, it does not follow that Joe knows that John's brother has red hair. This follows only if Joe also knows that James *is* the brother of John.

Now since the object of an intentional verb is only an object of the verb under the description and not under other descriptions that may be true of the object in question, it follows as Borst (1970) has pointed out that the property of being known, believed, wanted or looked for by someone is a property which the very same thing under another description need not possess; and hence that Leibniz's Law is not infringed, if the same thing is known, believed, wanted or looked for under one description, not under another description. Borst recognises the application of this argument in meeting such objections to the mind-brain identity theory as the objection that pain sensations cannot be brain processes, because many people who

have never heard of the brain and its processes, know that they have pains when they do. What he does not point out however, is its application in the case of the privacy of sensations. The reason for this perhaps, is that what is at issue in the case of privacy is not what a man knows about his experience as compared with what he knows about his brain processes, but how he comes to know what he knows in the two cases. This difference may make the application of the principle of nonsubstitutivity less obvious in the case of privacy; but it does not invalidate it. If, as I do, I know what my experiences are in a way no one else does, it not only does not but does not need to follow from the fact that my experiences are brain processes that I know what some of my brain processes are in a way that no one else does.

The Spatial Location of Experiences

The phenomenal properties of experience are properties of experience which allegedly cannot be properties of brain processes. The privacy of private experience is a case where experiences and brain processes each have a property which the other does not have, that of coming to be known about in two radically different ways. Spatial location and extension by contrast is a property which has a literal application in the case of all brain processes whereas, so the argument runs, spatial location and extension apply to private experiences in so far as they apply to them at all, only in a metaphorical sense. Hence by Leibniz's Law experiences and brain processes cannot be one and the same thing.

The first point that needs to be made in answering this objection is that the fact that we do not ordinarily locate our private experiences within our brains is no evidence that they are not in fact so located. For as we have seen, any micro-reductive account of an entity or phenomenon must necessarily attribute to the entity or phenomenon properties which the macro-description does not ascribe to it. Moreover, since 'knowing' is an intentional verb, the fact that we do not ordinarily know that our experiences are physically located within our brains, cannot be used as evidence that the two are not one and the same. What has to be shown is, either that experiences are not in the brain because they are somewhere else, or that experiences are not the kind of thing to which it makes sense to ascribe spatial location and extension.

The argument that some private experiences at least, are not in the brain because they are somewhere else, gets some purchase in the case of the sensations that we feel in the various parts of the body together with such things as the ringing that we hear in our ears or the spots that we see before our eyes. But as Descartes realised, the phenomenon of the *phantom limb*, where sensations are felt in a part of the body that is no longer there to be felt in, shows quite clearly that this kind of spatial location is not a literal spatial location. It is simply a device for indicating the part of the body in which the stimulus which is exciting the sensation in question *appears* to be located. To say that I feel a pain or a tingle in my left toe does not entail that there actually is anything unusual going on in my

toe, let alone that the sensation is actually going on there. Consequently, although it is misleading and confusing to say that the pain in my toe is literally in my head, there is no genuine contradiction between these two predicates. Nor do we need to say that because we now know that the sensation is actually in our brains, we were wrong in supposing it to be in the toe. As I see it, it is true of such a sensation both that it is in the toe and that it is in the brain. There is no contradiction between these assertions any more than there is a contradiction involved in saying that she went home in a taxi and a flood of tears. Such sentences sound odd only because they involve a juxta position of the preposition 'in' when it is functioning in two different ways.

The doctrine that spatial location and extension has no application in the case of the mental has been almost universally accepted by philosophers since it was first put forward by Descartes. But Descartes' argument for this view depends upon his contention that the mind is a substance or independently existing thing. His argument is that the mind cannot be a spatially extended substance because it cannot be chopped up into bits in the way that all spatially extended substances can be chopped up, leaving a set of separate bits behind each of which is a spatially extended substance in its own right. This argument is conclusive however, only if we accept Descartes' premise that the mind is substance. Once we abandon the notion that the mind is a substance and replace it with the notion of the mind as a system of interacting processes, events and states, we can happily agree with Descartes that the mind is not a spatially extended substance, but not because it is not extended, but because it is not a substance. Biological processes like the circulation of the blood are extended and located in space, but they cannot be chopped into bits each of which continues to exist as a process in its own right. One can cut off the blood supply from a part of the body and the blood will continue to circulate in the remainder; but once it is separated from the main system circulation inevitably ceases in the part that has been cut off.

Descartes' argument is not however, the only consideration that has helped to persuade philosophers of the truth of his contention that the mental cannot be said to be spatially extended and located. For there is undoubtedly something very odd about the suggestion that mental states like knowing, believing, wanting or intending are spatially extended or located. Why this should be is readily understood if we accept a Rylean hypothetical analysis of these mental dispositions. If what we are talking about, when we describe someone as knowing, believing, wanting or intending something, is not about anything that he is doing now or about anything that is currently the case, but only about what he would say and do if certain contingencies were to arise, we are not talking about anything which can be assigned any kind of spatial location, since the events we are talking about are hypothetical events which have not yet happened and may indeed never happen. Furthermore, if they do happen, they may happen in a large number of different places according to whether the individual happens to be at the time and, since they may consist in a variety of overt

acts of moving, manipulating and vocalising they cannot be sensibly located in any one part of the body.

If therefore, Descartes' thesis is understood as a thesis about mental dispositions, it is clear that he is quite right in claiming that such states are neither spatially located nor spatially extended. But the mind-brain identity theory in the form in which it was put forward by Feigl, Smart and Place was not intended to cover mental states. Its application was confined to the purely covert aspects of mental processes and in the case of Feigl and Smart to the raw or uninterpreted aspects of private experience. Smart indeed confined his argument solely to sensations: Now experiences and sensations are processes, things that involve continuous change which is extended over time. Moreover a process, in contrast to a dispositioned state, *is* the kind of thing which cannot be said to exist or to be going on unless it is going on *somewhere*. To say that something is not going on anywhere is tantamount to saying that it isn't going on at all. But since experiences in the relevant sense are clearly processes, it follows not merely that it makes perfectly good sense to assign them to a specific spatial location, but also that it makes no sense to assert that an experience is occurring, unless it is occurring somewhere. It is true that we cannot assign our experiences to any very specific location on the basis of introspective observation apart from the observation that visual experiences seem to take place somewhere behind our eyes and that the auditory experiences appear to happen somewhere in the head between the two ears. But apart from the confusion due to the two idioms of spatial location in the case of bodily sensations, there is nothing about our experiences, as we describe them from an introspective standpoint, which is inconsistent with the suggestion that they occur in whatever part of the brain the corresponding brain processes occur. Moreover it is arguable that the mind-brain identity hypothesis is the only theory which is capable of making sense of the logical fact that experiences are processes. For it is only on this theory that experiences can be given the precise spatial location which I would argue, they must have in order to make sense of the assertion that they actually occur.

IV. Central State Materialism

[I would like to thank Marco Nani for his help in preparing this paper. - U.T.P.]

Editor's Notes

<1> Jack Smart disagrees about this: "I think (as for example Quine does) that in mathematics 'equals' means 'is identical with'. Thus the number 7 plus 5 is identical to the number 12. Also the length in centimetres of this stick may be the very same number as the length in centimetres of that stick. You might say that two armies are equal but you could say that the number of soldiers in one is identical with (is the very same as) the number of troops in the other. 9 is identical to the number of the planets." (personal correspondence)

