



**OXFORD JOURNALS**  
OXFORD UNIVERSITY PRESS

## Mind Association

---

### Personal Identity

Author(s): H. P. Grice

Source: *Mind*, New Series, Vol. 50, No. 200 (Oct., 1941), pp. 330-350

Published by: Oxford University Press on behalf of the Mind Association

Stable URL: <http://www.jstor.org/stable/2250897>

Accessed: 20/05/2010 14:19

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=oup>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



Oxford University Press and Mind Association are collaborating with JSTOR to digitize, preserve and extend access to *Mind*.

<http://www.jstor.org>

## II.—PERSONAL IDENTITY.

BY H. P. GRICE.

I PROPOSE to discuss first the nature of the main question which philosophers have been asking, when they have concerned themselves with the problem of Personal Identity. Then I shall ask whether it is possible to maintain a Pure Ego theory of the Self; and finally I shall state and attempt to defend a form of Logical Construction theory. In fulfilling the first part of my programme I shall try to state rather dogmatically what I think to be the question really at issue between philosophers, irrespective of whether such philosophers would admit that this is the question or would agree with my formulation of it. I shall hope that the later sections of my article may provide some justification for my views about the nature of the question.

### *A.—The Question.*<sup>1</sup>

If we reflect on sentences in which the word "I" (or "me", etc.) occurs, we can, I think, distinguish at least three different classes of sentences, in each of which the use of the word "I" is different.

(1) Sentences such as "I am hearing a noise", "I am thinking about the immortality of the soul".

(2) Sentences such as "I played cricket yesterday", "I shall be fighting soon".

(3) Sentences such as "I was hit by a golf-ball", "I fell down the cellar steps".

Now in the sentences of my class (3) I can substitute, for the word "I", the words "my body" without loss or change of meaning. If I tell you that my body was hit by a golf-ball, I tell you neither more nor less than if I tell you that I was hit by a golf-ball; moreover, my use of words in the former case is quite a natural one, though perhaps less frequent than that in the latter. But in the case of sentences of my classes (1) and (2), no

<sup>1</sup> In what I say in this section, and elsewhere, I am under considerable obligation to Mr. Gallie's article "Is the Self a Substance?" *MIND*, 1936.

such substitution is possible. It is clearly unnatural to say "my body is hearing a noise"; and I think it is unnatural to say "my body played cricket" or "my body will be fighting soon". But though sentences of classes (1) and (2) have it in common that no such substitution can be made in them, they do, I think, differ from one another in respect of the use of the word "I" in them. I am inclined to think that the difference consists in the fact that any sentence of my class (2) is analysable, at some stage of analysis, into a sentence or sentences belonging to my class (3), together with a sentence or sentences belonging to my class (1). Thus the sentence "I played cricket yesterday" is analysable into a sentence or sentences stating something about the sort of movements I made (where "I" = "my body"), together with a sentence or sentences stating something about the sort of thoughts and intentions and decisions I had (where "I" does not equal "my body"). Of course these sentences into which sentences of my class (2) are analysable may well be themselves further analysable. The conclusion I draw, then, is that "I" sentences are of at least three kinds: one in which "I" can be replaced by "my body" without loss or change of meaning (3); another in which the sentence as a whole is equivalent to a sentence or sentences of the previous kind together with a sentence or sentences containing "I" used in a different sense from the previous sense (2); and, finally, sentences containing "I" used in this different sense first referred to (1).

Now I think it has been with sentences of the last mentioned kind (1), or with what has been stated by such sentences that most philosophers have been concerned, when dealing with the problem of Personal Identity. And I think that one way of putting the question that most of them have been endeavouring to answer, whether they have been aware of it or not, is to say that they have been trying to answer the question "What is the analysis of sentences of this kind?" Now I think this is not a bad way of putting the question, but I think there is a still better way. For every sentence of my class (1) there will be another sentence differing from the first in that where the first contains the word "I", the second contains the word "someone" (e.g. "I heard a noise": "someone heard a noise"). I will call these sentences "class (1) 'someone' sentences". (There will I think be as many different senses of "someone" as there are different senses of "I".) Now I think the clearest thing to say is that there are really two questions at issue: (a) What is the analysis of class (1) "someone" sentences? (b) What is the analysis of class (1) "I" sentences? (These two questions will

of course be connected : and the sort of answer I give to one may affect the question what sort of answer I can give to the other.) I think this clearer for two reasons : (1) it enables one to make clearer the relations between what Broad calls "The Proper Name Theory", "The Disguised Description Theory", and "The Logical Construction Theory". The P.N.T. and the D.D.T. both give one and the same answer to question (a): ("What is the analysis of class (1) 'someone' sentences?"), while the Logical Construction theory gives a different answer to question (a). But the P.N.T. and the D.D.T. differ from one another in the answers they give to question (b): ("What is the analysis of class (1) 'I' sentences?"); (2) it reveals a paradox in an asymmetrical view about the analysis of sentences about other people. For the asymmetricalist will have to maintain that any class (1) 'someone' sentence, *e.g.* "someone heard a noise" is ambiguous, or else that it is equivalent to a disjunctive sentence of the form ' $p$  or  $q$ ' where  $p$  is true when the speaker heard a noise, and  $q$  when someone else heard a noise : both alternatives being at the least surprising. I shall then maintain that the questions at issue are (a) What is the analysis of class (1) "someone" sentences? (b) What is the analysis of class (1) "I" sentences; and I hope the next section of my paper will do something to justify this contention.

### *B.—The Pure Ego Theory.*

People who have held or discussed forms of what, following Broad, I shall call a Pure Ego theory, have not usually formulated their statements as if they constituted answers to the questions I have enumerated. I propose to take as a sample the statement that "The Self is a Substance". The meaning of this statement has been discussed by Mr. Gallie, and I shall base what I have to say on his account. He suggests that people who have said that the Self is a Substance have meant what he would mean by saying that the Self is an ultimate particular (together with the assertion that the Self is the subject of mental but not of physical attributes, which I shall ignore, as irrelevant to my present purpose); and to say that the self is an ultimate particular is to say that the Self "has qualities and stands in relations, without either being or containing qualities and relations". This definition of "ultimate particular" is taken by Mr. Gallie to exclude from the class of ultimate particulars "all entities which are complex in the way in which the fact that 'This is red' or the event consisting in 'That noise being heard' are complex";

for such entities, though they may be particulars, contain qualities or relations as elements, and are not, therefore, ultimate particulars. In this way Mr. Gallie is able to maintain that the assertion that the Self is a Substance is an assertion which is *really* about the properties of things, and not about the properties of symbols.

I cannot regard this as a satisfactory formulation of a P.E. theory. The words "contain" and "element" are not defined, and I do not think it is possible to give them a sense which will allow to be true *all* of the things Mr. Gallie would want to maintain. In particular, whatever sense we give to "contain", I don't think Mr. Gallie is entitled to maintain *both* that it is impossible for something to be both an ultimate particular and a logical construction (which I am sure he would want to maintain) *and* that the proposition "The Self is a Substance" is *really* about the properties of things. Suppose first that "contain" bears the sense Mr. Gallie seems explicitly to attribute to it, and that X may be said to contain Y when X has to Y the relation which the fact "*This is red*" has to *red* or *redness* or the event *this noise being heard* to *being heard*. Then the Self will be an ultimate particular, even if it is a Logical Construction; for the sense of "contain" in which a Self, if it is a logical construction, could be said to contain, say, a relation which holds between two experiences of that self, or a quality of an experience of that self, will be a very different sense of "contain" from that in which the fact that *this is red* contains *redness*; and if this is so it is possible that the Self might be both a logical construction and an ultimate particular.

Suppose, on the other hand, we give "contain" the only sense which could possibly claim to fit Mr. Gallie's use of the word, and say that "X contains Y" is to be defined in the following kind of way "X has  $R_1$  to Y (where  $R_1$  = the relation which the fact *this is red* has to *redness*) or . . . or X has  $R_n$  to Y (when  $R_n$  is the relation which a logical construction has to some quality or relation)". But what can it mean to say "X has  $R_n$  to Y"? It seems to me it can only mean: X is a logical construction out of things of a certain kind, and one of these things has a relation R to a quality or relation Y (such as the relation which holds between the event *this noise being heard* and *being heard*). But to say that something is a logical construction out of something else is to assert a proposition about words: therefore a proposition of the form "X contains Y" is at least in part *really* about words; and therefore a proposition of the form "X is an ultimate particular" is at least in part about

words ; moreover, it is verbal in just that part of itself in which we are going to be interested if we are considering whether the Self is an ultimate particular.

What I wish to suggest is that either of two things may be meant by the assertion that the Self is a Substance. (1) To say the Self is a substance is to say some such thing as that selves persist and are capable of change and (perhaps) have causal properties, and so forth ; and even in saying this sort of thing I think we shall be asserting something about the use of words ; for something will be said about the way in which words like " someone ", " I ", etc., can be used significantly. For instance, " selves persist " might mean something like " If it makes sense to say ' someone has  $\phi$  ' then it makes sense to say ' someone has  $\phi$  both at  $t_1$  and  $t_2$  ' ". (I don't make any claims on behalf of this analysis of " selves persist " ; it is only intended as an illustration of what is probably the sort of thing asserted by such a proposition.) In this sense of the sentence, " The Self is a Substance ", the Self might be both a substance and a logical construction. (2) " The Self is a Substance " may mean (1) plus " the Self is not a logical construction " ; and to say that the Self is not a logical construction is to say that sentences such as " someone is hearing a noise " (*i.e.* class (1) " someone " sentences) are unanalysable in a certain respect. This will not mean that the sentence as a whole is unanalysable ; " hearing a noise " might be definable even if a P.E. theory is true ; and " someone " may mean " some person ", and " person ", too, may be definable even on a P.E. theory. What the P.E. theory does assert, I think, is that no final analysis of the sentence " someone is hearing a noise " can be given, which does not contain a variable (*e.g.* ' someone ' or ' something '), such that any proper name or description which can be significantly substituted for " someone " in the sentence " someone is hearing a noise ", can be significantly substituted for the variable in the sentence which constitutes a final analysis of the sentence " someone is hearing a noise ". For instance, if the final analysis of " someone is hearing a noise " were " something is a mental event and is related by R to the hearing of a noise ", then what the P.E. theory asserts about the analysis of " someone is hearing a noise " would not be true ; for in the sentence " someone is hearing a noise " I can significantly substitute the word " I " for the word " someone " ; but in the sentence " something is a mental event and is related by R to the hearing of a noise " I cannot substitute the word " I " for the word " something " without reducing the sentence to nonsense ; so the sentence " someone is hearing a noise " would not be

unanalysable in the respect in which the P.E. theory asserts that it is unanalysable. For brevity I shall describe the sort of unanalysability which the P.E. theory attributes to class (1) "someone" sentences by saying that what the P.E. theory asserts is that there is no new level analysis of class (1) "someone" sentences in respect of "someone". And what the L.C.T. asserts is that there is a new level analysis of class (1) "someone" sentences in respect of "someone".

It will I hope be seen that if the P.E. theory is formulated in this way, it avoids some of the objections which have been brought against it. For instance it will not be a valid objection to say that the P.E. theory introduces metaphysical entities, *i.e.* substrata. To say that substrata are metaphysical entities is to say, I suppose, that the proposition "there are substrata" is unverifiable. But if the P.E. theory is formulated in my way, either it does not follow, given the truth of the P.E. theory *plus* the truth of the proposition "there are selves", that there are substrata, or else, if it does follow, then to assert that there are substrata is to assert (1) that there are selves or people, (2) a proposition about the use of words, which may not be verifiable, but then, why should it be?

A much more serious objection to which the P.E. theory has been exposed is the following. Anyone who maintains a P.E. theory about the analysis of class (1) "someone" sentences, will also have to give an answer to my question (2), namely, "What is the analysis of class (1) "I" sentences?" Now it will be open to him to say either that when I use the symbol "I" in such sentences I use it as a logically proper name, in which case I shall be an object of acquaintance to myself; or that when I use the symbol "I" I use it as the equivalent of a descriptive phrase, in which case presumably, though not necessarily, I shall not be an object of acquaintance to myself. Now the argument against the P.E. theory will be that I am not acquainted with myself, so the P.N. theory of the analysis of "I" sentences cannot be right. But if the P.E. theory were right, the D.D.T. of the analysis of "I" sentences could only be right if I were in fact sometimes acquainted with myself. But I am not acquainted with myself: therefore if the P.E. theory were right neither the P.N. theory nor the D.D.T. could be right. But one must be. Therefore the P.E. theory is not true. I must now produce the arguments for this contention.

First, it may be said: That I am not acquainted with myself is shown by Hume. "When I enter most intimately into what I call myself, I always stumble on some particular perception or

other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe anything but the perception." To put the point more conveniently for the argument, we must admit that when we introspect, we are acquainted with experiences of various kinds, but never with anything other than experiences : and so with nothing of which " I " could be the proper name or description.

On the other hand, it will be urged, if we try to combine the P.E. theory with the D.D.T., we must allow that I am sometimes acquainted with myself. What sort of descriptive phrase will it be with which " I " will be synonymous in a class (1) " I " sentence ? I don't think it matters much for the purposes of the argument what it is, so I will assume that it is the phrase " the self owning this experience ", where " this " is a logically proper name. Thus " I heard a noise " will mean " the self owning this experience heard a noise " ; and, in general, any class (1) " I " sentence will assert that the self which owns one experience owns another. But here two observations become relevant. (1) We couldn't possibly know that any experience was owned by any self unless we were acquainted with the self that owned it, just as we couldn't know that anything was red unless we were acquainted with something which was red. We couldn't therefore know any facts such as that someone heard a noise unless we were acquainted with selves. But we do in fact know many such facts ; therefore we are acquainted with at least one self. (2) Even if objection (1) is not valid, and we could know, *e.g.*, that someone heard a noise without being acquainted with any self that did hear a noise, we still could not know that two different experiences were owned by the same self unless we were acquainted with a self which owned them both. But many class (1) " I " sentences state that two experiences are owned by the same self ; therefore unless we were acquainted with at least one self we could not know what is stated by any of these class (1) " I " sentences. But these class (1) " I " sentences include first those sentences which do state the sort of things we all of us from time to time know, such as " I heard a noise ", " I am seeing a red patch ". Therefore we have acquaintance with at least our own self, if the P.E. theory and D.D.T. are both true. But we have not acquaintance with our own self (shown above). Therefore not both the P.E. theory and the D.D.T. are true. But since the P.N.T. is false (shown above) the D.D.T. is true. Therefore the P.E. theory is false.

If, on the other hand, we reject the P.E. theory, it is easy to

see how both the D.D.T. and the proposition that we sometimes know the truth of what is stated by class (1) "I" sentences such as "I heard a noise", may be true. If, for example, to say that a self owns two experiences is to say that a relation of such and such a kind holds between the experiences, the relation might be of such a kind that I could sometimes know that it held between two experiences; and if so, then I might sometimes know such things as that I heard a noise.

The argument which I have just expounded may seem strong, but I do not think it refutes the P.E. theory. I do not propose to question the second part of the argument, the part that concerns the D.D.T. It is the first part of the argument which seems to me unsound, namely, the contention that when we introspect we are acquainted with experiences and nothing but experiences; and therefore I am not an object of acquaintance to myself. The fault in this part of the argument seems to me due to an over-carefree use of the word "acquaintance"; for I think "acquaintance" must be definable. (I am only concerned with "acquaintance" in the sense in which I can be said to be acquainted with particulars, and not in any sense in which I can be said to be acquainted with universals. Indeed, I have always found it difficult to see how to start to answer the question, "Am I acquainted with universals?"; and I think my difficulty is due to the fact that "acquaintance" is a technical term which has been given a use only in sentences which state something about acquaintance with particulars.) My reasons for thinking that "acquaintance", as used by philosophers, must be definable are: (1) It is a technical term; in ordinary life I just don't say "I am acquainted with a loud noise" or "I am acquainted with a thought of dough-nuts", or, for that matter, "I am acquainted with myself". In fact the only people with whom I am acquainted in any ordinary sense of "acquaintance" are people other than myself; but very few philosophers think that, in the philosopher's sense of "acquaintance", I am acquainted with other people. Not only is "acquaintance" in this sense not a word in ordinary use, but there is no word synonymous with it which is in ordinary use; "awareness" is the only candidate: and that has far too wide a meaning; for instance, there are circumstances in which I can say quite truly that I am aware of a table, but (according to most philosophers) I cannot even say truly that I am acquainted with a table. (2) The apparent occurrence of both acquaintance with particulars and knowledge of facts about them seems to invite the attention of Occam's razor; and it is much easier to define "acquaintance"

in terms of "knowledge of facts" than "knowledge of facts" in terms of "acquaintance".

So I suggest that "I am acquainted with X" (where X is a particular) means "(a) I know some fact about X, (b) X is not a logical construction". Or, I think, I can put the same suggestion another way, which will not involve the somewhat vague phrase "some fact about X"; namely, "I am acquainted with X" means "(a) I know some fact expressible in a sentence S containing 'this' (or some other demonstrative word) where 'this' refers to what 'X' refers to, (b) S is unanalysable in respect of 'this' (or the other demonstrative word in question)". (I put in (b) because a demonstrative word is not always used as a logically proper name; that is, it is often equivalent to a descriptive phrase.)

Let us now apply the first form of this definition to the sentence "I am not acquainted with myself"; this will mean "Either it is false that I know some fact about myself, or it is true that I am a logical construction". But I cannot *both* (a) deny that I know some fact about myself, *and* (b) claim that the P.E. theory and D.D.T., if both true, presuppose that I am acquainted with myself, on the grounds that if I were not I could not know facts about myself which I in fact do know. If I do I shall be claiming both to know and not to know facts about myself. So if I want to make the claim referred to in (b) above, the only ground on which I can assert that I am not acquainted with myself will be that I am a logical construction. But if this is so, the argument against the P.E. theory uses the proposition "I am a logical construction" as a premiss in order to show that the P.E. theory is false, that is, that the self is a logical construction. But no holder of the P.E. theory would be converted by such an argument. A similar conclusion will follow if I use the second form of definition of "acquaintance".

So far, I think, the P.E. theory is untouched. Can it then be refuted? I think it is not easy, but I will suggest one or two difficulties in it. The first difficulty is not easy to state shortly, but I will do my best. It seems clear that there are occasions when it would be true to say "someone is not now having an experience". Now this proposition must be distinguished from "it is not the case that someone is now having an experience" (*viz.* "no one is now having an experience"); and in order to do this, in the P.E. theory, I think we must say that "someone is not now having an experience" contains as part of its meaning "someone has now some characteristic  $\phi$ ". Now what sort of characteristic could  $\phi$  be?

(1) It might be some non-dispositional, non-relational characteristic. But it seems to me that the only non-dispositional, non-relational characteristics of selves with which we are familiar are characteristics consisting in the having of such and such an experience. But  $\phi$  could not be any such characteristic without making the proposition "someone is not now having an experience" self-contradictory.  $\phi$  will, then, have to be some characteristic with which we are not familiar; and this seems to me highly objectionable: for in order to have, as I'm sure I do sometimes have, evidence justifying a belief in the proposition "someone is not now having an experience" I should have to have evidence for the proposition "someone has  $\phi$  now"; but if I have no idea what  $\phi$  is, how could I? So it does not seem as if  $\phi$  can be a non-dispositional, non-relational characteristic.

(2)  $\phi$  might be some non-dispositional, relational characteristic. But this seems unsatisfactory: for then I think it would be a proper question to ask "What (or what sort of thing) is it which has  $\phi$ "; and I don't think we should think we had received a proper answer until we had been told some non-relational characteristic of the thing having  $\phi$ ; and this could only be a non-dispositional, non-relational characteristic (and so we get back to (1)), or

(3) a dispositional characteristic. It doesn't matter to my argument what dispositional characteristic  $\phi$  might be, so let me take for the sake of the argument the characteristic "capacity for thinking". It seems quite clear to me that when I assert "X is now capable of thinking", part at least of what I am asserting is a hypothetical proposition about X to the effect that X would now be thinking if so and so were the case (this proposition not being intended to exclude the possibility that X is now thinking). So I think "someone is now capable of thinking" must mean one of two things. It might mean (a) "someone would now be thinking, if so and so were the case". But this won't do; for suppose Adam had existed, but Eve had not: then it might be true that someone would now be thinking, if so and so were the case; for it might be true that Adam would now be thinking, if (*inter alia*) he were still alive. But since Adam would have died childless, it would not be true that someone is now capable of thinking. Or (b) "someone is now capable of thinking" might mean "someone has now some characteristic  $\psi$ , and would now be thinking if so and so were the case". But we now have to start all over again asking the same question about  $\psi$  as I have just been asking about  $\phi$ ; and we either have to say that  $\psi$  is a non-dispositional characteristic, which

is open to all the objections which I brought in (1) and (2) above against  $\phi$ 's being a non-dispositional characteristic ; or that  $\psi$  is a dispositional characteristic, in which case there will be yet another characteristic  $\chi$  which I assert something to have when I assert it to have  $\psi$ , and the trouble begins yet again.

I do not see any way out of this difficulty if the P.E. theory is true ; but if a Logical Construction theory is true the difficulty may not arise. For to say " someone is not now having an experience " may be to say something like " there have occurred and/or will occur some experiences having relation R to one another, and there would be now occurring an experience having R to each of these experiences, if certain conditions were realised ; but no such experience is now occurring ". And to say " It is not the case that someone is now having an experience " may be to say something like " No experience is now occurring which has relation R to any other experience ". (Of course the forms of analysis I have just given may not fit all types of L.C.T.)

My second difficulty is roughly this. Suppose the P.E. theory to be true ; and suppose I know that I had a headache yesterday, and that I had a toothache this morning. Now suppose that I am asked how I know that it is one self which had both experiences, and not two exactly similar selves. On the P.E. theory plus the P.N. theory, I don't see that I could give any true answer, except " I just do know ". This is, I think, rather unsatisfactory. But on a L.C.T., on the other hand, if I am asked this question, I can answer truly " Because the experiences have to one another the relation R which constitutes ' belonging to the same self as ' ". For instance I should answer " Because I remember (or know to have occurred) both experiences, and any experiences I remember (or know to have occurred) must be co-personal ". This answer would imply, I think, that the self is a logical construction, and is to be defined in terms of memory.

These objections are the only ones I can find against the P.E. theory, so I will now pass on to my third section.

### *C.—An Alternative Theory.*

The theory which I am going to suggest is, I think, mainly a modification of Locke's theory of Personal Identity. Exactly what Locke's answer to my first question (*i.e.* (a) What is the analysis of my class (1) " someone " sentences) would have been is not clear ; but I think it would have been that, for example, " someone heard a noise " means " the hearing of a noise (in

the past) is the object of some consciousness"; and "someone heard a noise and smelt a smell" means "the (past) hearing of a noise and the (past) smelling of a smell are objects of the same consciousness". This, I think, is borne out by Locke's words: "As far as any intelligent being can repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness as it has of any present action, so far it is the same personal self". (Consciousness, at any rate officially, for Locke means "consciousness of . . . as one's own".)

To this theory the following objections may be made.

(1) It is circular in so far as it defines  $x$  belonging to a self in terms of "consciousness of  $x$  as belonging to a self."

(2) Reid's puzzle about the officer, who was beaten for robbing an orchard as a boy, captured a standard when a young officer, and became a general; when he captured the standard he was conscious of having been beaten as a boy; when he became a general he was conscious of having captured a standard, but not of having been beaten as a boy. Therefore, according to Locke, the person who became a general was the same person as the person who captured a standard, and the person who captured a standard was the same person as the person who was beaten as a boy; but the general and the person who was beaten as a boy were not the same person. But this is absurd.

(3) If "consciousness of" involves knowledge of, very few experiences separated in time by a long interval could be personal (because memories are short).

(4) It is circular in a different respect from (1), in so far as it seems impossible to define "same consciousness" except in terms of "consciousness of (=belonging to) one person".

Difficulty (1) can be avoided by interpreting "consciousness" as meaning "memory", or "memory or introspection". The other difficulties require much more fundamental modification of the theory. This I shall undertake by stages, as my theory is somewhat complicated.

First of all I propose to introduce as a technical term the phrase "total temporary state". This term I shall define later; but I can indicate what I am talking about when I use the term by saying that a total temporary state is composed of all the experiences any one person is having at any given time. Thus, if I am now thinking of Hitler and feeling a pain, and having no other experiences, there will be occurring now a total temporary state containing as elements a thought of Hitler and a feeling of pain. Now since total temporary states may be

said to occur at various times, they may be said to form temporal series. (Such series may of course contain gaps : there may be times at which no member of a series is occurring, though members have occurred before these times and will occur after them.) What we want to do is to find something which will be true of any series of total temporary states all the members of which are total temporary states of one and the same person ; but false of any series of t.t.s., *not* all the members of which are t.t.s's of one and the same person.

As a preliminary shot I suggest the following : in a series of total temporary states belonging to one person, every t.t.s. which is a member of that series will contain as an element a memory of some experience which is an element in the temporally preceding member of the series ; in a series of total temporary states not belonging to one person this will not be the case. We can now give a provisional analysis of a class (1) " someone " sentence, *e.g.* " someone heard a noise ". This will mean " a (past) hearing of a noise is an element in a t.t.s. which is a member of a series of t.t.s's such that every member of the series contains as an element a memory of some experience which is an element in the preceding member ". This analysis will avoid the difficulties to which Locke's analysis is subject ; the paradox about the officer will not arise, nor will the objection that few remotely past experiences could be co-personal with present experiences ; and as far as I can see the analysis will not be circular unless it proves impossible to define " t.t.s. " except in terms of " person " or " someone ". But of that later.

Nevertheless, I do not think the analysis will do as it stands. It seems to me an unwarrantably violent assumption that every t.t.s. of mine (except the first) contains as an element a memory of some immediately preceding experience, indeed of any experience at all ; and that every t.t.s. contains at least one experience which is remembered immediately subsequently, indeed remembered at all. So I must amend the analysis to meet this objection. I propose to reconstruct it in terms not merely of actual memory but also of possible memory. The analysis of " someone heard a noise " will now run " a (past) hearing of a noise is an element in a t.t.s., which is a member of a series of t.t.s's, such that every member of the series *would, given certain conditions, contain* as an element a memory of some experience which is an element in the preceding member ". (For brevity I use " *x would, given certain conditions, contain y* " in such a way that it is true if *x* does contain *y*.)

A little further emendation is required, I think, to reach

a satisfactory analysis. For, first of all, there is a difficulty about the first t.t.s. of the kind of series I have been considering ; since it is the first t.t.s., there can be no preceding t.t.s. in the series ; so the first t.t.s. would not, given *any* conditions, contain a memory of a previous experience. Second, I don't think we should assume that every t.t.s. contains some experience which would, given certain conditions, be remembered in the immediately subsequent t.t.s. ; it might, for instance, be the case that even a prolonged process of psycho-analysis would not bring about the occurrence of a memory of any experience contained in some t.t.s.'s. To avoid these difficulties I will restate the analysis of "someone heard a noise" thus: "a (past) hearing of a noise is an element in a t.t.s. which is a member of a series of t.t.s.'s such that every member of the series *either* would, given certain conditions, contain as an element a memory of some experience which is an element in some previous member, *or* contains as an element some experience a memory of which would, given certain conditions, occur as an element in some subsequent member ; there being no subset of members which is independent of all the rest". (By denying that there is, within such a series, a subset of members which is independent of all the set, I mean to assert that any subset of t.t.s.'s includes at least one t.t.s. which *either* would, given conditions, contain as an element a memory of some experience contained as an element in some t.t.s. which is not included in the subset, *or* contains as an element some experience a memory of which would, given certain conditions, occur as an element in some t.t.s. not included in the subset. This proviso is obviously necessary in order to prevent the t.t.s.'s of a man who dies at *t*, and of another whose first experience occurs at *t*, from being by definition all t.t.s. of one person.) I can put the analysis more briefly if I introduce the term "memorative t.t.s." to mean "t.t.s. which would, given certain conditions, contain as an element a memory of some experience contained in a previous t.t.s.", and the term "memorable t.t.s." to mean "t.t.s. which contains as an element some experience, a memory of which would, given certain conditions, occur as an element in some subsequent t.t.s.", and the term "interlocking series" to mean "a series in which no subset of members is independent of all the rest" (in the sense of "independent of" I have just defined). Then "someone heard a noise" can be analysed "a (past) hearing of a noise is an element in a member of an interlocking series of memorative and memorable t.t.s.'s".

It now remains for me to define "total temporary state". "A t.t.s. occurs at *t*" means "experiences occur at *t* which

belong to the same t.t.s." ; and " experiences E and E' belong to the same t.t.s." means " E and E' would, given certain conditions, be known, by memory or introspection, to be simultaneous ". (I use " simultaneous " to mean whatever would be meant in ordinary speech by " occurring at the same time ".)

One final point must be emphasised before I discuss arguments for and against the theory I have propounded. By " memory " I must be understood to mean what is often referred to as " memory-knowledge ". I cannot interpret " memory " as, *e.g.*, " true belief about the past ". For clearly I can have a true belief that such and such an experience occurred, without the experience having been my experience. I should have to substitute for " true belief about the past " " true belief about *my* past ", and then my analysis would contain an obvious circularity. For I should have to analyse " someone " sentences in terms of " true beliefs about someone's past ". I think I must further maintain that not merely memory-knowledge, but also memory-acquaintance is possible ; that is to say it must be possible, given certain conditions, to know not merely that such and such an experience occurred, but also that *that* experience occurred. I do not however propose to argue this point.

I must now consider what there is to be said for my theory.

(1) It is a form of logical construction theory ; and since there seem to me to be grounds for rejecting the P.E. theory and also for rejecting all other forms of logical construction theory which I have encountered, there seems to me reason at any rate to investigate the theory I have suggested.

(2) On my theory it will be possible for some propositions about selves to be known. For two experiences can be known to be co-personal, if, *e.g.*, it can be known that memories of them occur within the same t.t.s., *i.e.* if it can be known that two memories occur simultaneously. But this can be known. Thus the theory has an advantage over theories of the self which do not allow knowledge of propositions about selves ; and there are several such theories.

(3) The theory, if true, enables us to see why such a proposition as " One can only remember one's own experiences " is analytic, and analytic in a way which is not trivial, as it would be trivial if " memory " were to be defined in terms of " having knowledge of one's own past experiences ". For even if we were to define " memory " in this sort of way, we should still be left with a question about the proposition, " one can only have knowledge of one's own past experiences ", which seems to me a necessary proposition ; and on the theory I suggest it will be analytic.

(4) The theory will recommend itself to those who feel an absurdity in saying "there have been experiences of mine which I could never, given any circumstances, be aware of". For if there were such experiences, they could not be elements in a t.t.s., since "E is an element in t.t.s." is defined in terms of the possibility of knowledge of E. But if they were elements in no t.t.s. they would belong to no self.

I shall conclude by discussing some objections which might be brought against the theory I have suggested. First of all, it may be said, the analysis I have suggested of a sentence like "someone heard a noise" is much too complicated to be the right analysis. I am far too uncertain how far, if at all, the fact that a proposed analysis of an apparently simple sentence is complicated is a good reason for rejecting the analysis, to discuss this objection at length. I will confine myself to the observation that my analysis of "self"-sentences is probably far less complicated than would be the phenomenalist's analysis of any material object-sentence, if indeed a phenomenalist were ever to offer an analysis of such a sentence, and not merely tell us what sort of an analysis it would be if he did give it.

The second objection which I may encounter is that my analysis of class (1) "someone" sentences is circular in a way something like that in which it has been said that phenomenalist analyses of material object-sentences are circular. To quote Braithwaite (*Propositions about Material Objects*, P.A.S., 1937-38, p. 275): "Now the most serious criticism to which such a theory (*i.e.* Phenomenalism) lays itself open is that the analysis proposed is circular: it is impossible to state the conditions under which a person will have a sense-datum of a clock on the mantelpiece without specifying a lot of things about the position of the person's body, the integrity of his visual and central nervous system, that he is not dreaming nor hypnotised—in fact a set of propositions which are, I think, equivalent to what I have called the perception in question being reliable." Now it might be urged that a similar criticism can be directed against my theory: for I have analysed "someone" sentences in terms of memories which would occur given certain conditions; but, it may be said, the conditions would have to include the occurrence of certain experiences other than the memories in question; and for the occurrence of such experiences to lead to the occurrence of the memories, it would be necessary that they should be experiences of the person to whom the memories would belong. But this involves a circle.

My answer to this objection is that my analysis of "someone"

sentences does not have to say what the conditions would be, given which a memory would occur : while a phenomenalist analysis of material-object-sentences may have to state what the conditions would be, given which a person will have a sense-datum belonging to a material object. If a phenomenalist analysis does have to state what the conditions are, one reason why it does have to do so may be the following : suppose I am in a room which does not contain a clock ; now the analysis of the sentence " There is a clock in the room " cannot be of the form " There are conditions given which I (or other people) would have sense-data of such and such a kind ". For if there is another room exactly similar to the first except that it does contain a clock : it will be true that there are conditions given which I (or other people) would have sense-data of such and such a kind (*i.e.* the kind I would have if there were a clock in the first room) ; for I should have such sense-data if (*inter alia*) I were in the second room. But it would still be false that there is a clock in the room (where the room meant is the first room). Now once the Phenomenalist has to state the conditions he is faced with the difficulty mentioned by Braithwaite. But I do not have to state what the conditions are given which a memory would occur ; for an experience, a memory of which would, given certain conditions, occur as an element in some t.t.s., belongs to the self of which that t.t.s. is a t.t.s., *whatever* the conditions in question are. Consequently, I think, my analysis is not open to a charge of circularity on this score.

The last possible objection to my theory which I shall consider is that my theory presupposes the occurrence of memory-knowledge ; but memory-knowledge never occurs ; so my theory is false. Strictly speaking, this is not an accurate way of putting the objection ; what actually is the case is that if my theory is true, and if any proposition expressible in class (1) " someone " sentences is true, then it is presupposed that memory-knowledge is *causally possible*, *i.e.* would occur given certain conditions. But I do not think my theory would be in the least plausible if memory-knowledge never did in fact occur ; and if my theory is true it certainly would not be possible ever to *know* that anyone had an experience, unless memory-knowledge sometimes occurred. So I think I am really committed to maintaining that memory-knowledge does occur.

Now most of the objections to the occurrence of memory-knowledge which seem to me at all serious seem to be serious only because they are directed against views concerning the nature of memory which maintain very odd things about the

mental images which are reputed to occur in memory situations ; they maintain, for instance, that in a memory-knowledge situation there is a mental image which is identical with a past event. But if we refrain from saying such odd things about mental images, and maintain that the function of a mental image in a memory-knowledge situation, if it has any function at all, is, to use Professor Price's word, merely "directive", we escape these objections. And, indeed, it seems to me perfectly clear that when I have memory-knowledge of something, it is not a mental image which I know, or about which I know something, nor does the proposition "I remembered something" entail the proposition "I had a mental image".

If this sort of objection is ruled out, what can the opponent of memory-knowledge say? He may just say that he has never had memory-knowledge of anything. If so I cannot really argue with him, I can only ask him whether he claims to know that he has never had memory-knowledge, or only to believe it ; and if he says he knows, ask him how he knows except by means of memory-knowledge ; or if he says he believes, ask him what his evidence is and how he acquired it.

But he might produce some further argument against the view that memory-knowledge does occur. Now the only argument I can think of which seems to me at all formidable is a causal argument, which might be stated thus. Suppose that memory-knowings do occur ; then, being events, they must be caused. What then is the cause? One view might be, the past experience, which is remembered, together with a stimulus which immediately preceded the remembering. But this involves the possibility of causation at a distance, which (it will be said) is very difficult to maintain. Another view might involve a persistent mental trace (the formation of which was caused by the past experience which is remembered) together with the present stimulus. But again, it will be said, the notion of a "mental trace" is a very difficult one. We are left then with the possibility that it is a persistent physical trace, caused by the past experience, in the body of the person who remembers, together with the present stimulus. Since this trace is usually supposed to be in the brain, I shall refer to it as a "brain-trace". For lack of an alternative, then, we must accept the view that the knowing is caused by existence of the brain-trace plus the occurrence of the stimulus. Now it is possible that the formation of the brain-trace might be caused, not by the past experience, but by, say, an operation by a clever surgeon. If this is so, it is possible that a brain-trace, exactly like that which would be produced by a past experience

of such and such a kind, might exist without any such experience having occurred. It will further be possible that both the brain-trace might exist and the stimulus might occur, without the past experience having occurred. But if both the brain-trace existed and the stimulus occurred, the memory-knowing would occur. Therefore the memory-knowing might occur without the remembered experience having occurred. But that is logically impossible. Therefore unless the argument is unsound one of the premisses must be rejected ; and the easiest premiss to reject is that memory-knowledge occurs.

Now I think the argument is unsound ; but in order to show that it is I must distinguish more closely what the argument asserts, for I think there is an ambiguity in it, due to an ambiguity in the word " possible ", which may mean either " logically possible " or " causally possible ". Suppose, first, that " possible " means " logically possible ". Then the bare bones of the argument will be :

(1) The existence of a brain-trace of kind A *plus* the occurrence of a stimulus of kind B is logically compatible with the non-occurrence of any experience of kind E.

(2) The existence of a brain-trace of kind A *plus* the occurrence of a stimulus of kind B causally involves the occurrence of a memory-knowing of an experience of kind E.

Therefore the occurrence of a memory-knowing of an experience of kind E is logically compatible with the non-occurrence of any experience of kind E.

But this is absurd ; therefore either (1) is false, which is very, very improbable ; or (2) is false, and the falsity of (2) will involve the falsity of the proposition that if these memory-knowings occur they are caused by the existence of a brain-trace *plus* the occurrence of a stimulus ; or there are no memory-knowings, which seems the easiest alternative to accept.

But there is a suppressed premiss in the argument which is false. (Perhaps it is rather a principle than a premiss.) The argument should run :

(1) The existence of a brain-trace of kind A *plus* the occurrence of a stimulus of kind B is logically compatible with the non-occurrence of any experience of kind E.

(2) The existence of a brain-trace of kind A *plus* the occurrence of a stimulus of kind B causally involves the occurrence of a memory-knowing of an experience of kind E.

(3) For any propositions  $p$ ,  $q$ ,  $r$ , if  $p$  is logically compatible with  $q$ , and  $p$  causally implies  $r$ , then  $r$  is logically compatible with  $q$ .

Therefore the occurrence of a memory-knowing of an experience of kind E is logically compatible with the non-occurrence of any experience of kind E.

But (3) only has to be considered to be seen to be false. Let  $p$  = it has been raining,  $q$  = the ground is not wet,  $r$  = the ground is wet. Then  $p$  will be logically compatible with  $q$ , for it is logically possible that it should have been raining without the ground being wet; and  $p$  will causally imply  $r$ , for whenever it rains the ground does get wet; but  $q$  is clearly not logically compatible with  $r$ ; for it cannot be true both that the ground is wet and that it is not wet.

I conclude then that the argument in this form is unsound; but before I pass on to the second form the argument might take, I ought to remark that it must not be supposed that I accept the views about the causes of memory-knowledge involved by the argument.

Suppose now that "possible" means "causally possible". The argument (including the suppressed premiss or principle) will now run:

(1) The existence of a brain-trace of kind A *plus* the occurrence of a stimulus of kind B is causally compatible with the non-occurrence of any experience of kind E.

(2) The existence of a brain-trace of kind A *plus* the occurrence of a stimulus of kind B causally involves the occurrence of a memory-knowing of an experience of kind E.

(3) For any  $p, q, r$ , if  $p$  is causally compatible with  $q$ , and  $p$  causally implies  $r$ , then  $r$  is logically compatible with  $q$ .

Therefore the occurrence of a memory-knowing of an experience of kind E is logically compatible with the non-occurrence of any experience of kind E.

But this is absurd; therefore (as before) we must reject memory-knowledge.

(3) is now, I think, true; but its gain is (1)'s loss. For the supporter of the argument is now committed to maintaining not that it is logically possible that a brain-trace of kind A should exist without the occurrence of an experience of kind E, but that it is causally possible that it should so exist. That means, I think, that he has got to maintain that there are conditions given which there *would* be a brain-trace of kind E without any experience of kind E having occurred; and in order to support this contention he must maintain, for example, that if a surgeon operated in a certain way he *would* produce the brain-trace, or give some other explanation how the brain-trace could be produced. But to maintain any such thing as this is something,

I should have thought, that no reasonable man would be prepared to do. For I cannot see what evidence in favour of it *he* could possibly have.

I do not then think that any real doubt has been cast on the occurrence of memory-knowledge ; and it seems to me, therefore, that my theory is untouched by objections of the kind I have just discussed.