

their own pasts. Since it is a conceptual truth that memory statements are generally true, it is a conceptual truth that persons are capable of knowing their own pasts in a special way, a way that does not involve the use of criteria of personal identity, and it is a conceptual truth (or a logical fact) that the memory claims that a person makes can be used by others as grounds for statements about the past history of that person. This, I think, is the kernel of truth that is embodied in the view that personal identity can be defined in terms of memory.

NOTES

*This article is reprinted from *The Journal of Philosophy*, Vol. 56, No. 22 (October 22, 1959), by courtesy of the editor and Professor Shoemaker.

1. Thomas Reid, *Essays on the Intellectual Powers of Man*, ed. by A. D. Woozley (London: Macmillan, 1941), p. 206.

2. H. P. Grice, "Personal Identity," *Mind*, (October, 1941), 340.

3. I use "remember" in its most common sense, in which "I remember that P" entails "P," and "I remember X occurring" entails "X occurred."

4. John Locke, *An Essay Concerning Human Understanding*, I, ed. by Fraser (Oxford: The Clarendon Press, 1894), 457.

5. Roughly speaking, a statement is a memory statement if (supposing it to be an honest assertion) it cannot be false unless the speaker has misremembered. A conclusion from what is remembered, on the other hand, can be false without there being a mistaken memory. E.g., I mistakenly identify the man I saw as John when in fact it was his identical twin.

6. The word "generally" is vague, but I doubt if this can be made much more precise. This statement should perhaps be qualified so as to apply only to memory beliefs concerning the *recent* past.

7. We can, of course, have inductive grounds for believing that one person's memory claims are exceptionally reliable and that another's are exceptionally unreliable.

8. Ludwig Wittgenstein, *Philosophical Investigations* (Oxford: Basil Blackwell, 1953), I, paras. 56 and 265.

In John Perry (ed.): *Personal Identity*.
University of California Press,
1975, 135-155.

John Perry

9

Personal Identity, Memory, and the Problem of Circularity

When it is asked wherein personal identity consists, the answer should be . . . that all attempts to define would but perplex it."¹ When he said this, Butler was thinking of Locke's² attempt to define personal identity in terms of memory; if his opinion about a future state, which motivated his interest in personal identity, proved correct, he has doubtless since had similar thoughts about more recent "memory theorists," such as H. P. Grice³ and Anthony Quinton.⁴ For in spite of such perceptive critics as Butler and Reid,⁵ the thought that personal identity is analyzable, and analyzable in terms of memory, has been periodically revived.

In this essay, I try to discover the strengths and weaknesses of the memory theory, by defending the best version of it against arguments that could be raised by those, who feel as Butler did that the concept of personal identity is primitive. The memory theory emerges from this defense with its letter intact but its spirit scathed.

GRICE'S THEORY

Locke suggested that *A* is the same person as *B* if and only if *A* can remember having an experience of *B*'s.⁶ The sufficient condition implied is plausible: if I really can remember going to the store yesterday, then I must have gone to the store. That is, I must be the same person as someone who went to the store. But the implied necessary condition is much too strong,

as Reid and other critics have pointed out. That I cannot remember going to the store yesterday does not mean that I did not go. Forgetting, even beyond the possibility of recall, is possible.

Later memory theorists have concentrated on weakening the necessary condition to the point of plausibility. Grice, whose account is, in my opinion, the most subtle and successful, in essence takes Locke's relation, disjoins it with its converse, and takes the ancestral of the result. Grice adopts the notion of a total temporary state, or t.t.s., which is a set of simultaneous experiences of a single person, and conceives of his task as finding the relation that must obtain between t.t.s.'s that belong to one person. In Grice's terms, with *A* and *B* now being t.t.s.'s and not persons, the relation Locke uses in his analysis is this:

R_L : *A* contains, or would contain given certain conditions, a memory of an experience contained in *B*.

The relation that results from Grice's weakening maneuvers we can express this way:

R_G : There is a sequence of t.t.s.'s (not necessarily in the order they occur in time, and not excluding repetitions), the first of which is *A* and the last of which is *B*, such that each t.t.s. in the sequence either (i) contains, or would contain given certain conditions, a memory of an experience contained in the next, or (ii) contains an experience of which the next contains a memory, or would contain a memory given certain conditions.⁷

A set of t.t.s.'s which can be formed into a sequence of this sort, and to which no more t.t.s.'s can be added (which I shall call a "Grice-set"), is a person or self.

Grice's account avoids the Brave Officer Paradox⁸ and other stock counterexamples to memory theories of personal identity, to which his predecessors and successors have fallen prey. But it is not at all obvious that he avoids objections of another sort, in the spirit of Butler's criticism of Locke: "memory presupposes, and so cannot constitute, personal identity."⁹ In this essay I will examine three charges of circularity, each maintaining for a different reason that Grice implicitly uses the concept of personal identity in his analysis of it.

I shall not examine every interesting objection of this sort that could be made against Grice; in particular, I shall not ex-

amine the objection that experiences themselves, the ultimate building blocks in Grice's constructions of persons, must be individuated in terms of persons. I do not believe this objection is fatal, but discussion of it would lead us away from the topics I wish to discuss, into the difficult problem of the individuation of events.

CIRCLES AND LOGICAL CONSTRUCTIONS

Before settling down to specifics, we must satisfy ourselves that Grice's enterprise is of a sort for which circularity is a vice. He explicitly defends the view that persons are logical constructions from experiences. Whether he held this view as a part of a generally phenomenalist philosophy is not disclosed in the article on personal identity, and Grice may well have had special views about the nature of the logical constructor's enterprise, and special motivation for holding persons to be so constructable. But it will be helpful and only fair, given lack of contrary evidence, to suppose Grice involved in a logical construction of a "standard" sort.

The logical constructor attempts to analyze sentences about objects of some category, into sentences about objects of some other category. Examples of such analyzed and analyzing categories are numbers and classes, material objects and sense-data, persons and t.t.s.'s. The analyzing sentences may themselves be thought analyzable—for example, sentences about classes into sentences about propositional functions, or sentences about t.t.s.'s into sentences about experiences. At the bottom of the structure are sentences with a favored epistemological status, as, for example, that they can be directly known, because the objects they are about can be directly inspected. Through analysis, this favored status, or at least some status more favorable than was originally apparent, is transmitted up the structure to the analyzed sentences. Talk about persons might have seemed to involve us in talk about pure egos, or substances of some other obscure sort, but when we see that talk of persons is, really, just talk of t.t.s.'s and ultimately, of experiences, our knowledge is revealed as more secure than it seemed.

Sentences about experiences seem to be directly knowable, by the people who have the experiences, at the time they occur. Now a present tense sentence about persons, or material objects, cannot plausibly be regarded as merely a remark about present experiences. But it has been thought that they could be plausibly regarded as asserting no more than would be asserted by a string of sentences about past, present, and future experiences. While not all these sentences could be directly known at one time, each of them could be directly known at some time. The complex of assertions, into which a sentence about persons or material objects can be analyzed, will not have as favored an epistemological status as a sentence about a present experience. But it will have a more favored status than a sentence that asserted things never directly knowable by anyone at any time.

(In both the construction of material objects and of persons, it soon becomes clear that past, present, and future experiences do not provide sufficient materials: possible experiences, the experiences someone would have had, had things been different than they were, are also needed. And it is not clear that sentences about possible experiences have much favored epistemological status to transmit upward.)

On Grice's conception of logical constructions, if all goes well, the analyzed sentence (say, "Someone heard a noise") and the analyzing sentence (say, "A past hearing of a noise is contained in a t.t.s. which is a member of a Grice set") will have just the same truth conditions.¹⁰ If this were the only condition of a successful analysis, the analyzed sentence could serve as the analysis of the analyzing sentence, for "has the same truth conditions" expresses a symmetrical relation. It is the favored status of the analyzing sentence which gives the logical construction its noncircular structure.

A charge of circularity against Grice, then, will consist of two claims. First, that the analyzing sentence does not seem to have the favored status, and so must itself be analyzed. Second, that its analysis will have to employ sentences about objects of the category constructed, that is, sentences about persons. This would show that even if Grice has produced an analysis free from counterexample, it is a failure: the mystery of personal identity is transmitted downward to memory, rather than the

clarity of memory being transmitted upward to personal identity.

THREE CHARGES OF CIRCULARITY

The core of Grice's analysis is R_L . R_L is in itself a disjunction; the first charge of circularity will concern the first, simpler, disjunct:

A contains a memory of an experience contained in *B*.
The second and third counts of circularity concern the second disjunct:

A would contain, given certain conditions, a memory of an experience contained in *B*.

For simplicity, I explain these charges in an assertive tone, but the reader should keep in mind that ultimately I shall reject them.

(i) Smith examines a green cube, and later vividly describes his examination of it. Jones has never examined a green cube; he is hypnotized, and told that when he awakes he will remember examining one. Jones later vividly describes examining a green cube. To observers who do not know the whole story, Smith and Jones both seem to be remembering, in vivid detail, a past examination of a green cube.

Smith is really remembering, Jones is not. Their present experiences, the occurrence of which they know directly through introspection, are indiscernible. Jones cannot discover he is mistaken through careful attention to his own mind. Their outward behavior, the sentences they use, their facial expression, etc., is also indiscernible. And yet Smith's experience is a memory of a past experience, and Jones's is not. Saying of an experience, that it is a memory, is thus a complex attribution, and not just a report of what is directly observed through introspection. That a person is really remembering at a given time, and not just seeming to, cannot always be determined solely on the basis of observations of the person made at that time, whether by that person or others.

That we are *seeming* to remember a past experience, can be known directly; that we are really remembering involves more. And this more is not just the occurrence of a past experience of the appropriate sort directly knowable when it occurred.

For there was a past experience of the sort Jones *seems* to remember—Smith's experience of examining the cube. What further must be added? The example suggests that one further necessary condition is that the same person who is seeming to remember have had, in the past, the experience in question. But then, spelled out, with the full analysis of memory incorporated into the condition, the first disjunct of R_L would look like this:

A contains an apparent memory of an experience contained in B, and A is a t.t.s. of the same person of whom B is a t.t.s., and . . .

The ". . ." represents whatever further conditions may be found necessary for an analysis of memory. But we need go no further. The italicized condition is sufficient to doom Grice's analysis to circularity.

(ii) Even if the last objection is somehow overcome, Grice's analysis would still be circular, in virtue of the subjunctive conditional, "would, given certain conditions," contained in the second disjunct of R_L .

Let us look at the kind of example that makes this disjunct necessary. Wilson is asleep. His present t.t.s. contains only a vague blissful feeling, which he will never remember after awakening. Thus there are no actual memory links between Wilson's present t.t.s. and his past (because his present t.t.s. contains no memories), and there never will be any actual memory links between Wilson's future t.t.s.'s and his present one. So the analysis cannot rely solely on the first disjunct of R_L .

Had we shaken Wilson a moment ago and asked, "What thrilling things did you do today?" he would now be telling us about seeing Wynn hit a home run at Dodger Stadium earlier in the day. Although his t.t.s. contains no memory of this past experience, given certain conditions (our having shaken him and asked him the question) it would now contain such memories. It contains, we might say, only *possible* memories of the past t.t.s. The second disjunct asserts that there will always be at least a chain of possible memory links where there is personal identity.

Now, the problem with this conditional is not simply that its truth cannot be known through any sort of direct observa-

tion, but that, taken literally, the sentence "t.t.s. A would, under certain circumstances, contain a memory of seeing Wynn hit a home run," makes no sense. To make sense of it we will have to use the concept of personal identity.

It makes no sense, taken literally, because the identity of a t.t.s. must be determined by the experiences it contains. A t.t.s. is a set of experiences, and a set's whole identity is wrapped up in its membership.¹¹ The t.t.s. or set of experiences Smith would have had, if he had been awakened and questioned, and the t.t.s. he actually has, while asleep, are different t.t.s.'s. When we say, "The t.t.s. would have contained a memory . . ." we can only mean something like "The person would have had a different t.t.s. than he did have, and that different t.t.s. would have contained a memory." And in making sense of the conditional, we have had to talk about persons.

An analogy may help to make this point clear. When we say, "If the meeting had been advertised, the number of people in the hall would have been greater," we don't mean to imply that there is a certain number, say 50, which would have been greater if the meeting had been advertised. The number 50 will always be a little greater than 49 and a little less than 51, no matter how well advertised meetings are. Rather, we mean that a different number, say 101, would have fit the description, "number of people in the room," had the advertising been more thorough. So with the t.t.s. Wilson had and the t.t.s. Wilson would have had. They are not the same t.t.s., but different t.t.s.'s, one which deserves, and one which would have deserved, the description, "Wilson's t.t.s."

In order to state a conditional like the one about the meeting or the one about Wilson fully and explicitly, we need some "anchor"—some entity that retains its identity under the imagined change in circumstances, and in terms of which the number or t.t.s. is identified. In the case of the meeting, the meeting itself is the anchor: the *same* meeting would have drawn a different number of people. And in the case of Wilson, Wilson himself seems the natural anchor: the *same* person would have had a t.t.s. that contained a memory . . . had he been awakened and questioned. But then, fully spelled out, the second disjunct of R_L is:

Given certain conditions, the same person of whom *A* is the t.t.s. would have had a t.t.s. that contained a memory of an experience of *B*'s.

But this uses the concept of personal identity, and so the analysis is circular.

(iii) Even if charges (i) and (ii) are somehow circumvented, the phrase "given certain conditions" leads to a third problem.

Should Grice tell us which conditions it is, under which t.t.s. *A* would contain a memory of an experience contained in t.t.s. *B*? If he simply means "There is at least one condition such that, if it obtained, t.t.s. *A* would contain memories of an experience contained in t.t.s. *B*" then he owes us no such list, the analysis is complete as it stands. But if not just any condition will do, he should tell us which ones will.

But it seems quite clear that Grice cannot mean simply "There is at least one condition such that . . ." by the phrase "under certain conditions." For if he does mean this, "Under certain conditions, t.t.s. *A* could contain a memory of an experience of t.t.s. *B*'s" would not mean anything like what it is supposed to mean, viz., "The person, of whom *A* is a t.t.s., can remember an experience of *B*'s." Consider this example. Johnson saw a flash of lightning in the sky last Thursday; immediately afterward he received a serious head injury. As a result he cannot remember seeing the flash—the injury, we may suppose, interfered with the consolidation of short term memory which makes memory of such events for more than a few seconds possible. In this case, we would not say "Johnson can remember seeing the flash of lightning." No amount of reminding or prompting will bring it about that Johnson remembers. But we can state a condition such that, if it had obtained, Johnson would now be remembering the flash of lightning: that he didn't receive an injury, and was just asked if he had ever seen lightning. (We may suppose Johnson had never seen lightning before, and would surely have remembered it, if not for the injury.) But the fact that the conditional, "If Johnson had not been injured, and had just been asked about it, he would now be remembering seeing the flash of lightning" is true, does not show that Johnson *can remember* seeing the lightning, even though the truth of some other conditional like "If Johnson were not

asleep, and had just been asked about it, he would remember . . ." would show that he can remember seeing it. So some conditionals of the form, "If *C*, then Johnson would remember . . ." are relevant to the claim that he can remember, and some are not. So the words "t.t.s. *A* would, given certain conditions . . ." must mean "there are certain conditions, *C*₁, *C*₂ . . ., and under one of these conditions t.t.s. *A* would . . ." Grice owes us a list, or some other specification, of these conditions.

I wish to make, but not press here, the point that it is unlikely this could be done. The point essential to this charge is that, even if the conditions were exhaustively listed, it seems inevitable that the concept of personal identity would be required. The only example we have discussed so far of such a condition is that the person with t.t.s. *A* was awakened a few moments ago and questioned; if, under those conditions, the person would remember, then, under actual conditions, he *can* remember. Now it is hard to see how this condition, or any of the conditions involving prompting, reminding, threatening, all of which typically occur somewhat before the occurrence of the t.t.s. in question, could be expressed without requiring that it be the *same* person who is prompted, etc., who is later to remember. If the phrase "given certain condition" were cashed in, as it must be, for a list of conditions, the second disjunct of *R*_L would look like this:

t.t.s. *A* would contain, if *the same person* who has *A* had been awakened and asked, or if *the same person* who has *A* had not just taken a powerful drug, or . . . a memory of an experience contained in *B*.

And so, again, we see that Grice's analysis makes implicit use of the concept of personal identity and is circular.

MEMORY

Memory can be analyzed without use of the concept of personal identity, and Grice thus cleared of these charges of circularity. I sketch such an analysis here, focusing first on the ordinary way of expressing event memory, as in "MacKenzie remembers Wilbur's marriage," or "Sandy remembers seeing her high marks," and later considering Grice's rather specialized lo-

cution, "t.t.s. *A* contains a memory of an experience contained in t.t.s. *B*."

Event memory must be distinguished from memory *that* or factual memory, particularly from factual memory that an event occurred. Most of us remember *that* Columbus discovered America in 1492. We wouldn't miss that question on an exam. But no one now alive remembers *Columbus discovering America*. Most of us remember that we were born; few of us remember our birth. We can remember *that* events occurred which we never witnessed, and no plausible account of personal identity could be built on factual memory. But we can only have memories of events that we witnessed or in which we consciously took part. (This last is added because of the peculiarity of saying that I "witnessed" the event of my drinking my coffee; I didn't witness that event, I was a part of it. Having noticed this peculiarity, I shall go on to use "witness" in this extended and peculiar way, to cover both witnessing of and participation in an event.) This last fact, which I shall call the Witnessing Condition, was appealed to in the discussion of Smith, Jones, and the green cube. The issue between the memory theorist and his critic is not whether the Witnessing Condition is true, but whether it is a part of the analysis of memory, as the critic maintains, or a consequence of the analysis of personal identity, as the memory theorist does.

Events are commonly designated by nominalizations of sentences: "John's hitting of Mary," "the sinking of the Titanic," etc. Event memory is commonly expressed by prefacing such an event designation with words of the form "*X* remembers." But such event designations do not usually completely identify the event that they are being used to designate. Thomson and Robinson might both truly say, "I remember Zimbalist's becoming confused," though they remember different events; Zimbalist is easily confused. Most event designations completely identify only an event *type*. I shall use the letter "*E*" as a dummy event designation and as a variable ranging over event types, and "*e*" as a variable ranging over specific events. "*X* remembers *E*" I take as making the claim that (i) there is an event *e* which *X* remembers, such that (ii) *e* is of type *E*. I rely on the intuitive feel for what it is for an event to be of a type, in order to concentrate on (i).¹²

The analysis of memory requires three sorts of conditions, having to do, in turn, with what must happen at the time of the remembering, what must have happened at the time of the remembered event, and what the link between the remembered event and the event of remembering must be.

The first condition I call the Representation Condition. Representation is a notion I borrow from Martin and Deutscher's excellent discussion of memory.¹³ It has been thought, for example by Locke,¹⁴ Hume,¹⁵ and Russell,¹⁶ that mental imagery is required for memory of an event. This is a mistake. Someone giving a vivid verbal description of a past event, or painting a picture of it, could be said to be remembering that event, whether or not he was having, or could produce, mental imagery of the event. But something separates the rememberer and the apparent rememberer from the common run of mankind. Martin and Deutscher introduced the term "represent" to cover the many ways a person can indicate the past occurrence of an event of a certain type, and I follow them not only in adopting this notion, but in apologizing for not giving a fuller account of it.

The first step in our analysis of "*A* remembers *e*," then, is (1) *A* represents the past occurrence of an event of some type *E*.

What sort of thing is *A*? *A* is to be a live human body, or a human being. The difference between this concept and that of a person has been emphasized by many writers on personal identity, and is a point of agreement between memory theorists and Butler and other critics who think personal identity an unanalyzable concept. So I shall feel free to use the concept of a live human body, and of bodily identity, in the analysis of memory, without fear of circularity.

Condition (1) is satisfied by both the real and apparent rememberer, as well as others who comment on the past: factual rememberers, liars, historians, and the like.

The second condition is a detoxified version of the Witnessing Condition:

(2) *B* witnessed event *e*.

I shall call this the Weak Witnessing Condition. It makes no claim of identity between *A* and *B*.

Now suppose we had added, as held necessary in the first charge of circularity, the (strong) Witnessing Condition. This would have disqualified Jones as a rememberer, but the analysis would still be deficient. Hennig examined the green cube, then received an electrical shock that wiped out his memory. The Electrical Company, in compensation, had him hypnotized, and given the same posthypnotic suggestion as Jones. Hennig satisfies (1) and the Witnessing Condition, but is not a rememberer. So, even if we had the Witnessing Condition in the analysis, we would still need a third condition, a Linking Condition, to rule out Hennig. It seems clear that what would be further required is some condition to the effect that the past witnessing *bring about* the present representing. My strategy, in what follows, is to beef up the Linking Condition in such a way that the Witnessing Condition is not needed.

With or without the Witnessing Condition, it is not easy to see what exactly the Linking Condition should be. I believe that the view Martin and Deutscher defend, that the link is a causal one, is correct.¹⁷ But, as they point out, merely requiring that if the witnessing had not occurred, the representing would not be occurring, will not do. If Hennig had not examined a green cube, the Electrical Company would not have underwritten his hypnosis, and he would not be representing. (And Smith, the rememberer, would be representing, even if he had not examined the cube, for in that case I would have had him hypnotized and treated like Jones.) The witnessing must not just cause the representing, it must cause it *in a certain way*.

Scientists are trying to discover the causal mechanisms involved in memory. Suppose they discover that a certain process is involved in memory. Could our linking condition simply be that *that* process led from *B*'s witnessing to *A*'s representing? No, for in analyzing the concept of memory we seek beliefs common to all who use with understanding the formula "*X* remembers *E*," and knowledge of, or even specific beliefs about, the processes involved in memory are not at all common.

But we may believe that memory involves some characteristic process, without having a belief about which process, or what kind of process, that might be. In fact, I think we do believe this. Some who have the concept of memory may be sure

the process is not, or not merely, a material one; this was apparently Bergson's view.¹⁸ Others may believe it certainly is a material process, an electrochemical process of the central nervous system. Perhaps most have no opinions on the matter. But in accepting, as we all do, that "He remembers it" is an explanation of representing; in predicting, as we all do, that in certain circumstances people are likely to remember the past, and in other circumstances unlikely to; in seeking, as we all do, alternative explanations for representing of the past when circumstances make memory unlikely ("He can't have remembered, he was too young—his mother must have told him."), we indicate that we do believe there are certain processes involved in memory, which can be expected to occur in some circumstances, and not in others. This is a hypothesis, a speculation if you will, for no such process can be observed by the ordinary man, introspectively or otherwise. But it is an irresistible hypothesis.

Let us say that a witnessing and a representing are *M*-related when they are the beginning and end of such a process. Then our analysis is simply:

A remembers *e* if and only if

- (1) *A* represents the past occurrence of an event of type *E*;
- (2) *B* witnessed *e*;
- (3) *B*'s witnessing of *e* is *M*-related to *A*'s representation of the past occurrence of an event of type *E*.

But is it fair to use, in the analysis, a relation the nature of which we haven't disclosed? It is fair only if we can identify the relation, independently of the concept analyzed. This I have not done, for all I have said about the *M*-relation is that it is the relation involved in memory. But I shall now try to provide such an independent identification of the *M*-relation.

"Recollection" I shall use purely as a technical term, for which I stipulate this definition:

A recollects *e* if and only if

- (1) *A* represents the past occurrence of an event of type *E*;
- (2) *B* witnessed *e*, and *e* is of type *E*;
- (3) *B* and *A* are the same live human body.

Recollection, so defined, occurs often. One of the things we all know about live human bodies is that they are quite likely to recollect, and we know the conditions that make recollection

more and less likely. But recollection is a significantly different notion from memory. Returning to the case of the green cube, both Smith and Hennig recollect examining the cube, though only Smith remembers. With regard to cases that actually occur, memory is a more restrictive concept than recollection. Oddly enough, with regard to cases produced in the imagination of man, memory seems less restrictive. Philosophers thinking about personal identity, seeing no contradiction in trans-bodily memory, have produced many characters who remember what they do not recollect: Locke's prince,¹⁹ Shoemaker's Brownson,²⁰ Quinton's no longer fat but still apolaustic Pole.²¹ And the occupants of the Hereafter are regularly conceived as remembering earthly events, although the "resurrected" bodies of those occupants must not be the very same bodies as were buried and rotted away on earth. So the concept of memory is not simply more restrictive and not simply less restrictive than recollection, but sits askew of it.

An *unaided* case of recollection is one in which the representing of *A* is not explained by provision of information about *e* other than *B*'s witnessing of it. Now any ordinary human is drawn to the belief that there is an explanation for the frequent occurrence of unaided cases of recollection, that there is some process, material or immaterial, gross or sublime, complex or simple, which frequently occurs when a human being witnesses an event and leads to that same human's later representation of it. When the witnessing of an event leads by this process to a later representation of it, the witnessing and the representation are *M*-related.

I now have identified the *M*-relation not just as the relation that links the witnessing and representing in memory, but, non-circularly, as the relation that explains the great bulk of cases of recollection. And, of course, it is not an accident that the *M*-relation plays both roles.

My view is that the key to understanding memory is seeing it as an explanatory concept, not merely in that individual cases of past-representing are explained by memory, but that a generalization about human behavior, the frequency of recollection, is explained by an hypothesized process, and that this process is incorporated into the very concept of memory. This concep-

tion of memory explains its skewed relation to recollection. Memory is a more restrictive concept, in that more is required, the witness and the representor must not just be the same human being, but a certain process must have occurred. But by distinguishing between the *M*-relation and the relation of being or belonging to the same human body, and by virtue of our lack of knowledge of the nature of the *M*-relation, it becomes possible to think of the two as separate; we are able to imagine the possibility that certain witnessings and representings might be *M*-related, though not experiences of the same human body. It does not follow, after all, from the fact that the *M*-relation is regularly associated with sameness of human body, that it must always be so associated. And indeed we can, through use of the *M*-relation, extend the class of remembers. We can let *A* and *B* in our analysis stand for, not just human bodies but human bodies and any other sorts of things, ghosts or even gorse-bushes, that might, for all we know, become *M*-related to them.

There is another dissimilarity between memory and recollection. In a case of recollection, the representation must be accurate, the event recollected must be of the type represented, but no such condition has been placed on memory. We do not require a person's memory of an event to be accurate. Smith may be rattling on about the time he met the Prince of Wales in London; Jones may quite correctly observe that Smith never met the Prince of Wales, and has never been in London, but is really remembering the time when, as a part of a hoax that defies summary, he met Stanky in Philadelphia. The point is not that Smith speaks truly when he says "I remember meeting the Prince of Wales in London." His claim, remember, is twofold, that he remembered a certain event, and that it has a certain type, that it was a meeting of a Prince of Wales in London. The point is rather that Jones speaks truly when he says Smith is remembering meeting Stanky in Philadelphia, even though Smith is not representing the past occurrence of an event of *that* type. The event remembered need not be of the type represented. This too is explained by the suggested relation between recollection and memory. We build the concept of memory on a relation, the *M*-relation, in which we are interested largely because it so often leads to accurate past-representing. But we allow that the

processes involved, when conditions are less than ideal, may not inevitably lead to accuracy.

If we add to the three conditions of memory these two:

(4) *e* is of type *E*;

(5) *A* believes (1)-(4);

we shall have what I call a *paradigm* case of memory. Paradigm cases explain our interest in memory as a source of knowledge about the past; only when a person is remembering accurately, and knows he is remembering, and not, say, imagining, can he derive knowledge of the past from his own tendency to represent it.

What is the relation between "*A* remembers *e*," the concept just analyzed, and "t.t.s. *A* contains a memory of an experience of t.t.s. *B*," the expression Grice uses? I take it that experiences are a species of events. But it will not do simply to say, as an explication of Grice's notion, "*A* remembers *e*, and *e* is an experience." For suppose Wilson remembers Wynn watching the ball go over the fence. Then Wilson is remembering an experience, but Wilson's present t.t.s. does not contain a memory of an experience contained in Wynn's earlier t.t.s., in Grice's intended sense, or else Grice's analysis is in more serious trouble than contemplated so far. The experience we are after is not the event remembered, even if it is an experience, but the witnessing of it. Now given our peculiar use of "witnessing," the witnessing may be the event remembered. Wynn remembers watching the ball go over the fence, and it is this very watching of the ball which, in virtue of our extended use of witnessing" as including participation in the past event, is, in his case, the witnessing of the remembered event. But when the witnessing and the event remembered are distinct, it is the witnessing, and not the event witnessed, that belongs in the rememberer's biography. So I shall take "t.t.s. *A* contains a memory of an experience contained in t.t.s. *B*" to mean "*A* is representing the past occurrence of an event of some type *E*, and this representing is *M*-related to *B*'s witnessing of some event *e*."

Now we must turn to the charges of circularity, to see if Grice has been cleared.

(i) This charge rested on the claim that the Witnessing Condition must be incorporated into the analysis of memory. But

I have argued that with a properly formulated Linking Condition, the Weak Witnessing Condition is sufficient. The Witnessing Condition is not rejected. It remains true, a consequence of the analysis of memory plus Grice's analysis of personal identity.

(ii) This charge was that in order to make sense of the conditional used in the expression of possible memory, we had to take the person as the "anchor," the entity that stayed the same under the imagined change of conditions. We could, I think, answer this by simply taking the human being involved to be the anchor. But in replying to the third charge, we shall eliminate the use of subjunctive conditional in the expression of possible memory, making the present charge irrelevant.

(iii) We certainly have a concept of possible memory, of persons who could remember a certain event, although they are not in fact doing so. And there are certainly conditions such that, if their obtaining would lead a person to remember, then it is true of him that he can remember. But it would be a mistake to approach the concept of possible memory by trying to list these conditions.

A better approach to the problem begins with the notion of an inclination to believe that an event of type *E* occurred. Someone who is inclined to believe that an event of type *E* occurred will be disposed to represent that such an event occurred at that time. We do not need to have an exhaustive list of the conditions under which this disposition will be triggered in order to understand what it is to be so disposed, any more than we need to have an exhaustive list of the conditions under which a belief will be expressed, in order to know what it is to believe. Now, just as we believe that humans often represent the occurrence of past events of a certain type as a result of a certain process set in motion by a past witnessing, we also, I think, believe that a person may have such a disposition to represent, as a result of such a process. Indeed, we believe that having such a disposition is a part of the process that eventually leads, in some cases, to representation. Thus we can introduce the *M'* relation, which obtains when the processes that lead from witnessings to dispositions to represent occur, and analyze *A* has a possible memory of *e* as follows:

- (1) *A* is disposed to represent the past occurrence of an event of type *E*;
- (2) *B* witnessed *e*;
- (3) *B*'s witnessing of *e* is *M'*-related to *A*'s being disposed to represent the past occurrence of an event of type *E*.

For this analysis to be legitimate, we should provide an independent identification of the *M'* relation; this could be done along the lines used before, by first constructing a notion of possible recollection, and introducing the *M'*-relation in terms of the processes that explain the frequency of unaided possible recollection.

LOGICAL CONSTRUCTIONS AND INFERRED ENTITIES

Although I have defended Grice against the charges of circularity, the concept of memory I have used does not fit well with his conception of a person as a logical construction from experiences.

If a person is a logical construction from experiences, the existence of a person should follow, as a matter of logic, from the occurrence of the experiences of which the person is composed. The existence of a person entails nothing more than the existence of those experiences, related in a certain way, a way that itself could be immediately read off from experience. Thus Russell contrasts logical constructions with "inferred entities," where the word "inference" carries the implication of a non-demonstrative inference, incorporating some element of probability, or some explanatory hypothesis, that goes beyond the directly known facts.²²

But when I say that my toothache this morning, and my headache of last night, belong to the same person, because my toothache belongs to the same t.t.s. as a memory of the headache, we are saying, according to the concept of memory just defended, that a certain process, the nature of which we do not know, led from the headache to this morning's toothache-accompanied memory impression. The occurrence of this process does not follow from the occurrence of the headache, the toothache, and the memory impression. The occurrence of the process, and so of the person who both had the headache and has the toothache, is in fact an inference, not something directly known at

all. We believe that there is such a process at all since that seems the most likely explanation of the frequency of recollection; we believe such a process was involved in this case because of a lack of alternative explanations, and because it seems very likely that such a process should have occurred, given the other things we believe, including things believed on the basis of memory; for example, that given last night's other activities, a headache was to be expected; that given last night's sleep, with no evidence of interruption by electrical shock, mad scientist, brain transplant, or hypnotist, a memory of it was to be expected.

Also in the explanation of possible memory, that which might be directly knowable was sacrificed for what can only be inferred. A memory impression, an "occurrent" belief, a representing, may perhaps be objects of direct observation, but beliefs in the ordinary sense, in which I have many beliefs with which my mind is not now occupied, a disposition to represent, a possible memory,—these are all states we ascribe to persons, including ourselves, as a way of systematizing and explaining the conditions under which more directly observable phenomena occur. Indeed, in using subjunctive conditionals in his formulation, Grice had already left the realm of what, in any reasonable sense, can be directly known.

So neither the primitiveness of memory nor the primitiveness of personal identity is suggested by our investigation, but only the derivative nature of both concepts. And they are derivative, not from the conception of a world of atomistic experiences, but from our scheme of a material world of which human beings are a part. And the nature of the derivation is not logical construction, but generalization and theory building in the service of explanation and prediction. And if such theories as the belief in a process that explains recollection lead us to speculations and even convictions that carry us well beyond the material world that forms their evidential base, that is a danger of the natural human bent for such theory building against which must be weighed its utility in the mundane tasks from which these speculations provide an occasional relief.

I end with two disclaimers. I do not think Grice's theory, even freed from its origins in the project of logical construction, and incorporating the concept of memory defended here, is fully

satisfactory. As Quinton saw, ways in which a person's past are expected to influence his future other than just event memory, should be incorporated into our account of personal identity. The pattern used in doing this, however, could be one suggested by our investigation of Grice, first elaborating generalizations about human behavior after the pattern of our concept of recollection, and then introducing the relation which is believed to underlie them, and forms the basis of our concept of a person. But this is a large project.

Secondly, the approach that has emerged from our investigation of Grice is not inimical to Locke's original scheme, for Locke was not a logical constructor and had a place, in his version of the memory theory, for unknown processes and inferred states. This fact has often been sighted as a sign of his faint-heartedness, in not banishing from his philosophy the last traces of the notion of substance, but I think it is rather a sign of his levelheadedness. And Locke would also, I think, be sympathetic with the point made above, for it is only by generalizing from the memory theory, and incorporating somehow into our account of personal identity the sort of character development, stability of ideals and values, influence of past intentions, and the like, which we normally expect to find in humans, that the forensic and moral importance of personal identity, which Locke so rightly emphasized, can be explained.

NOTES

1. Joseph Butler, "Of Personal Identity," originally an appendix to *The Analogy of Religion* (1736); reprinted in this anthology.
2. John Locke, "Of Identity and Diversity," chap. 27 of *An Essay Concerning Human Understanding*, 2d ed. (1694); reprinted in this anthology.
3. H. P. Grice, "Personal Identity," *Mind*, vol. 50 (October, 1941); reprinted in this anthology.
4. Anthony Quinton, "The Soul," *Journal of Philosophy*, vol. 59 (July, 1962); reprinted in this anthology.
5. Thomas Reid, "Of Mr. Locke's Account of Our Personal Identity," from chap. 6 of *Essays on the Intellectual Powers of Man* (1985); reprinted in this anthology.
6. Locke's actual words are, "as far as this consciousness can be ex-

tended backwards to any past action or thought, so far reaches the identity of that person . . ." ("Of Identity and Diversity," sect. 9).

7. Grice actually describes the series, co-membership in which is required of *A* and *B*, as follows, "every member of the series either would, given certain conditions, contain as an element a member 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." ("Personal Identity," sect. C.) The condition imposed on *A* and *B* by R_G is equivalent to this; a series of the sort Grice describes would result from the R_G sequence by putting the members in chronological order and eliminating repetitions; an R_G sequence can be obtained from a series of the sort Grice describes by starting with any t.t.s. and building an appropriately linked sequence, repeating a multiply linked t.t.s. when necessary in order to continue until all the t.t.s.'s are used. For a comparison of Grice, Quinton, and Locke, see the introduction to this anthology.

8. See Reid, "Of Mr. Locke's Account of Our Personal Identity."

9. Butler, "Of Personal Identity."

10. This seems to be a difference between Grice and Russell, the pioneer logical constructor. In Russell's view, the logical construction was the philosopher's contribution to an improved conception of, say, a material object, free of the epistemological problems inherent in the ordinary conception. So analysis, for Russell, does not preserve exact meaning. But Grice intends to be making explicit, through analysis, the concept we already have. (See Bertrand Russell, *Our Knowledge of the External World* [New York: W. W. Norton, 1929]).

11. Grice does not say explicitly that a t.t.s. is a set. But it seems clear that if a t.t.s. is not a set, it is nevertheless some other sort of entity the identity of which is determined by the experiences contained in it.

12. Our "intuitive feel," of course, won't take us far when the difficult cases of event individuation and classification arise.

13. C. B. Martin and Max Deutscher, "Remembering," *Philosophical Review* (1966).

14. John Locke, "On Retention," chap. 10 of Book II of his *Essay Concerning Human Understanding*.

15. David Hume, "Of the Ideas of Memory and Imagination," Sect. III of Pt I, Book I, *A Treatise of Human Nature*.

16. Bertrand Russell, "Memory," chap. 9 of *An Inquiry Into Meaning and Truth* (London: Allen and Unwin, 1921).

17. But I do not accept their final version of this condition.

18. Henri Bergson, *Matter and Memory* (London: Allen and Unwin, 1912).

19. See Locke, "Of Identity and Diversity," sect. 15.

20. Sydney Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca: Cornell University Press, 1963) p. 23.

21. See Quinton, "The Soul," sect. 3.

22. See *Mysticism and Logic* (New York, 1929), pp. 155 ff.