Chapter 11

Temporal consciousness and confabulation: Escape from unconscious explanatory idols

Gianfranco Dalla Barba

‘How did your disease begin?’ asks the doctor to patient CA

‘It started with a strong sore throat… I couldn’t swallow anything… so I couldn’t go to school… my mother called the doctor’, answers the patient.

‘How did your disease begin?’ asks the same doctor to another patient, CD

‘It started with a strong headache… one morning I woke up with a strong headache and then I started throwing up and I remember I couldn’t keep my eyes open’, answers the patient.

One of these two patients is confabulating, whereas the other one is reporting the memory of an event she has really experienced. But who is confabulating, CA or CD? As you can see, there is nothing in the patients’ reports that can help you to tell which patient is confabulating. However, if I tell you that CA is a 67-year-old woman with Korsakoff’s syndrome (Dalla Barba et al., 1990) and that CD is a 33-year-old woman reporting the onset of her Herpes meningitis, things become much more clear and these additional pieces of information allow you to identify CA as the confabulating patient. This is not much because you know that patients with Korsakoff’s syndrome confabulate, but rather because you know that it is quite unlikely that somebody who is 67-year-old goes to school and has a mother who calls the doctor for her sore throat. In addition you know that headache, vomit, and photophobia are common in the onset of Herpes meningitis, which suggests you that CD is not confabulating. Nevertheless, CA’s and CD’s reports have something in common. They are semantically appropriate (Dalla Barba, 1993), in the sense that a hypothetical observer not familiar with the patients’ history could hardly tell whether their reports are confabulatory or not. In other words, the semantic structure of CA’s and CD’s memories is internally consistent, but the content of CA’s
memory is externally inconsistent with the patient’s past and present situation. CA remembers ‘another’ past (Dalla Barba et al., 1997b), in the sense that she is aware of a personal past that is different from the past she would have remembered if she wouldn’t have had a severe amnesico-confabulatory syndrome. The case of patient MG (Dalla Barba et al., 1997a) well illustrates how semantically appropriate confabulations can go undetected when the real present and past situation of a patient is unknown. While he was waiting to undergo a CT scan, MG told the radiologist that he had accompanied a friend to be admitted to the neurology department that day. The neurologist who was taking care of MG’s (inexistent) friend realized that MG also had neurological problems and so decided to refer him to the radiology department for a CT scan. On that occasion the radiologist did not even suspect that MG was confabulating.

However, patients who confabulate like CA, not only remember ‘another’ past, but are also disoriented in time and place. They are living ‘another’ present, in the sense that they are present to a world which differs from the world to which an observer is present. Consider the following example.

‘How old are you?’ asks the doctor to patient PL, a 57-year-old woman, hospitalized following a heart arrest (Dalla Barba et al., 1998).

‘I’m 20’

Then the doctor asks her to look at herself in the mirror.

‘Well, I look 50-year-old, but it’s because I was sick. I fell in a ditch while I was playing with my brothers. That’s why I damaged my face.’

‘Where are you now?’

‘I’m in a school to learn.’

‘Why then do you wear a night gown?’

‘Well I was hospitalized… I was hospitalized somewhere… I don’t even know what the problem was. You should ask my teacher’

‘How many children do you have?’

‘Two’ (actually 3).

‘What are their names?’

‘Roberto and Susy’ (actually Lucia, Luisella and Georgia).

‘How old are they?’

Roberto is three and a half, Susy… I don’t know (her children were actually 36, 33, and 10).

So, what is PL’s present? It’s the present of a 20-year-old woman, with two little children, who is at school to learn. She is present to a world, which is in sharp contrast with the real world, that of a middle-age woman, with a young daughter and two other adult children, who is hospitalized for a severe medical
condition. However, the distorted world to which PL is present is a *real* world *for her*, like *real* is the remembered distorted past and the distorted present world for patients who confabulate. In fact, confabulators deeply adhere to their confabulatory world and tend to act upon their confabulations. Metcalf’s and co-workers’ patient (Metcalf et al., in press), for example, upon his return home from the hospital, would awake each morning with the conviction that he was required to attend a swimming carnival at school that day and would attempt to search for his swimming costume, despite being 36-year-old and unable to walk due to injuries sustained in the accident that provoked his confabulatory amnesia.

So far we have seen that confabulation involves the patient’s past, as revealed by confabulatory memories, and present, as shown by confabulatory reports and acts about the present. Is it that? Is confabulation limited to past and present? Certainly, it would be quite surprising that a patient who lives with distorted personal past and present would live without a personal future. In fact this is not the case since these patients also confabulate about their personal future. Patient MB (Dalla Barba, 1993), for example, while he was hospitalized, said on one occasion that he was looking forward to the end of the testing session because he had to go to the general store to buy some new clothes, since he hadn’t been able to the day before, because he had got lost in the center of Paris, where he had fortunately met a nurse who kindly took him back to the hospital. On that occasion the patient actually attempted to leave his hospital room, claiming that there was a taxi waiting for him downstairs. This is obviously not the only example of confabulation about personal future. When asked to report their plans or projects about their future, confabulators use to report as personal plans their habits and routines. Patient GA (Dalla Barba et al., 1997b), for example, to the question ‘What are you going to do in a few minutes?’ answered ‘I’ll go home to cook supper’, whereas since her disease she never cooked and she was living one and a half hours from the hospital she was admitted to. To the question ‘What are you going to do tomorrow?’, she replied ‘I’ll go out shopping alone by car’, whereas she never did that since her disease and she would certainly not do it in the following day.

Plausible verbal reports and acts involving the patient’s past, present, and future are not the only example of how confabulating patients can interact with their world. Actually, confabulation is sometimes described as a heterogeneous phenomenon. In fact, some patients tend to confabulate spontaneously, while others confabulate more when in response to direct questioning. The content of confabulation can also vary considerably, being in some cases frankly implausible. The main difference between these cases and cases in
which confabulation is plausible is that confabulation can be detected without knowing any further information about the patient. Consider the following example.

‘What did you do yesterday?’ asks the doctor to patient SD

‘Yesterday I won a running race and I have been awarded with a piece of meat which was put on my right knee’.

In order to tell that SD’s report is confabulatory, you don’t need to know that he is 37-year-old patient who reported a severe head trauma. In fact, SD’s confabulation describes an event that not only he is unlikely to have experienced at that particular time, but also it is unlikely that anybody would have or will ever experience. The content of SD’s confabulation is semantically anomalous, in the sense that it carries meanings, which are inconsistent with knowledge and information shared by the members of the society. Confabulations often described as fantastic or implausible actually contain semantic anomalies (Weinstein and Lyerly, 1968; Stuss et al., 1978; Baddeley and Wilson, 1986; Joseph, 1986; Sandson, Albert et al., 1986). However, although semantically anomalous, these confabulations are often made of autobiographical elements put together in an inappropriate semantic structure. Consider again SD’s example. SD was actually involved in running races and had spent most of his time off work in this kind of activity. It was actually during a running race in the mountains that he fell reporting a severe head trauma and an open wound in his right knee. So, his confabulation about having won a running race and having been awarded with a piece of meat on his knee actually contains autobiographical elements inserted in an inappropriate semantic context.

From the description of confabulating patients reported so far emerge three characteristics of confabulation: (1) confabulation affects memory, in particular, but not exclusively, episodic memory (Tulving, 1972, 1983), and manifests itself when patients retrieve personal episodes from their past. (2) The content of confabulation can vary considerably, ranging from semantically appropriate to semantically anomalous. (3) Confabulation is not only limited to the patients’ past, but also involves their present and their personal future.

Why confabulating patients make such errors when retrieving their past? Why confabulation is in some occasions indistinguishable from a true memory, whereas in other occasions has such a bizarre or semantically anomalous content? Is confabulation a pure memory disorder, or the fact that it involves the patient’s past, present, and future reflects a disruption of how personal temporality is experienced? In the following sections we will address these questions.
Confabulation affects memory and past. So, at first glance it seems that in confabulating patients there is something wrong in the way they retrieve their past. Are these patients retrieving the ‘wrong’ past or remembering ‘another’ past, as we said earlier? Or, putting it in a different terminology, are they activating the wrong memory trace, as proposed by some theories, e.g. (Schnider and Ptak, 1999)? In order to answer these questions, we need first of all to know what the past is. Although a presupposition about the nature of the past is necessarily implicit in any theory of memory, this presupposition has never been clarified. The being of the past is somehow negated in these theories, in the sense that the past is considered as what is no longer. This assumption has produced the theory of memory traces: if something is no longer present but continues to exist in the form of memory, it must be preserved in the form of a more or less stable modification in the organism which remembers. Accordingly, if I now perceive the image of this cup of coffee on the table, tomorrow I will be able to remember it because this image has, so to speak, been deposited in some part of my brain in the form of a memory trace. In other words, the event of ‘cup-on-the-table-on-such-and-such-an-afternoon’ has caused a modification in the equilibrium of my brain which I call memory trace of the event, or, to use more elegant terminology, an engram, in which the event is deposited in the form of representation. The activation of the trace, that is, its subsequent passage from a passive to an active state (as in copy theories of remembering) or its interaction with a present cue (as in reconstructive models of memory), will result in recollection of the event. Figure 11.1 depicts this point of view: event E is stored in memory as memory trace E′; the activation of E′, with or without the involvement of a retrieval cue, results in E″, i.e., the conscious remembering of E.

Present-day versions of this now classical way of thinking have been reformulated more or less explicitly in current functionalist theories. Functionalism maintains that mental operations are expressible in an abstract and physically neutral language that designates functions and functional relations (Marcel, 1988). As far as memory is concerned, functionalist theories substitute the physical trace with a functional trace which is preserved in the architecture of a system which is no longer physical (the brain), but virtual (the cognitive system). It is easy to see that, however reformulated, any theory which bases the ability to recollect on the preservation of an event inside a trace contains a paradox, fruit of a misleading assumption, that is the belief that time can exist in things. The past event which I now remember, for example my dinner yesterday with Paul, is present to my consciousness as a past event, in that as
such, that in being past, that event was contained in the memory trace whose activation produced the recollection. In other words, the past of that event, or its ‘pastness’ as in Bergson (1896), is already there, enclosed in that ‘thing’, physical or abstract as it may be, which I call memory trace of the event. But it should be clear that ‘things’ as such are not temporal. Objects of the world are neither present, nor past, nor future, but they acquire a temporal dimension only in the presence of a person who goes to the trouble of making them temporal. Time cannot even be found in a ‘thing’ like the brain, however much one would like to give it a particular status in the universe. This erroneous assumption, on which theories of memory are founded, is directly reflected in the paradox to which these theories fall victim. The paradox of the memory trace, as we shall call it, consists in seeing memory, that is, what is given to consciousness as past, as originating in elements borrowed from the present. Let’s see why.

The memory trace paradox

The event which I now perceive, for example, the glass which is on the table in front of me, is without doubt a present event. This event of glass-on-the-table determines a modification in the equilibrium of a system, be it physical (the nervous system) or virtual (the computational level), which I call memory trace. What is the temporal nature of this modification, namely of the memory trace? Without doubt present. The glass-on-the-table which I now perceive is present and if one accepts that this event produces a modification somewhere in my brain or in my mind, one will have to accept that said modification will
be present and that the event represented by that modification will also be present. In short, depending on the level of description that one wants to adopt, be it a case of the synthesis of new proteins, of the growth of new dendritic spines, of the activation or the reinforcement of synaptic circuits, or of the particular codification of information, it will always be a matter of present modification of something. What happens when that event contained in the trace is recollected in memory? When the event of 'glass-on-the-table' is recollected, it happens in the present, that is, as the result of the reactivation of the modification that the event caused on a physiological, biochemical, neuroanatomical, neurocibernetical or functional level. And so it is not at all clear how memory, whose basic characteristic is being memory of the past, can stem from a combination of present phenomena, perception, the preservation of the trace, and recollection, which do not reflect any notion of past. Figure 11.2 depicts this situation: event ‘E’ is present and stored as present ‘E’ in the memory trace. When E’ is retrieved this happens as a consequence of a present process. Including whether a present cue is involved or not, the result of this processing is ‘E”’, i.e. the conscious remembering of E as ‘past’.

Activation of the memory trace should, if anything, coincide with a new perception of the event contained within the trace, not with its memory since the event contained in the trace was present as it ended up in the trace and continued to be so for as long as it remained enclosed in the trace. The paradox of the memory trace, which current theories on memory overlook had already been clearly reported over 60 years ago. Merleau-Ponty (1945) wrote, ‘...our very best reason for rejecting physiological preservation of the past is also a
reason for rejecting 'psychological preservation', and this reason is that no preservation, no physiological or psychic 'trace' of the past can enable comprehension of consciousness of the past. This table carries traces of my past life, I have carved my initials, I have left ink stains. But, alone, these traces do not refer back to the past: they are present and, if I find there signs of some 'previous' event, I find them because, by other means, I have a sense of the past, because I carry in me this meaning. – a preserved perception (in a 'physical' or 'psychic' trace) is a perception, it continues to exist, it is always in the present, it does not open behind us that dimension of escape and of absence that is the past.' If, on the other hand, I recognize that particular event as past, this happens because I attribute a precise meaning to it, that of being past, a meaning which by definition cannot be contained in the trace since, in every moment of its existence, it has never ceased to be present.

This does not mean that events do not cause modifications in our brain or in our cognitive system, but these modifications cannot be used to explain recollection. Besides, any object in the universe undergoes modification on account of events. Any object carries with it the marks of the past, and yet no one would think of attributing the possibility of recollection to objects, if not metaphorically. The signs that events have left on objects acquire the meaning of past only by virtue of a consciousness which attributes it to them. Even the impression on the cushion on the sofa in front of me bears witness to past events, but in itself it is present and has never ceased to be so in any single instant of its existence. If I find in it signs of some past event, if I remember the thousands of times that my guests and I myself have sat on that sofa, it is because I attribute a meaning to those marks, that of being past. The 'meeness' of memory of which Claparède (1911) speaks, the 'warmth and intimacy' with which James (1890) describes memory, or the 'pastness' of Bergson (1896) describe certain characteristics of memory well, but tell us nothing about its nature. Indeed, they either describe only one present characteristic of memory, and so they remain in the present, or else if they are already dealing with the past, they presuppose that which they want to explain. Certainly, there are events that seem to be 'itself' past without belonging to the personal experience of any particular individual. For example, the big bang itself exists in the past, but no human experienced it. Would that allow to make a distinction between the past itself and a human past, or our concept of past? Certainly no. The very necessary condition to be satisfied for the big bang to be past is the existence of a present consciousness that considers it as past. Without such a consciousness, the big bang is neither past, nor present nor future: it just doesn’t exist. At the most, a distinction must be made between a generic impersonal past
and *my* past. When I say that I remember the big bang, I’m not remembering my personal past, I’m just conscious of a piece of information, a notion, that contributes to my knowledge of the world: the big bang is an hypothesis concerning the origin of the universe. We will analyze this in more detail when we will discuss the distinction between Knowing Consciousness and Temporal Consciousness.

**Tags, labels and the past**

In support of the hypothesis of the memory trace as a condition for recollection it could be stated that it is not at all necessary for episodes to be contained in the memory trace as past. Indeed, for recollection to be possible, it is enough for ‘something’ in the trace to indicate that its content concerns some past event. In short, it would be enough for the episodes registered in the trace to be in some way marked, for example, with a type of tag stating the date they took place or simply indicating that what is contained in that trace is past, e.g. Anderson and Bower (1972), (1974); Hintzman (1978); Morton et al., 1985. It is easy to see how yet again a hypothesis of this type presupposes the past instead of explaining it. Any tag or indicator added to the episode contained in the trace would indeed be an indicator in itself present just as the episode which it accompanies. The date or the ‘past’ information of the indicator, in themselves, tells us nothing, they are not past but they become so only if the past is already given for the consciousness which picks up that indicator. In other words, the label or indicator of pastness of an episode does not precede nor, even less so, does it create the past. In contrast, it is in some way the consequence of it, it assumes its function as indicator of pastness because I already carry in me this meaning and am capable of attributing it to it. The knot I tie in my handkerchief to remind me that yesterday I made an appointment with Paul is not past and if I ignore it, it isn’t even present. When I put my hand in my pocket to take out my handkerchief and realize it has a knot, I attribute a meaning to that knot, that of being an indicator of a past event which I have to remember so as not to cut a poor figure with Paul. If I didn’t already possess the past as meaning, that knot would simply remain what it is and not an indicator of something past. The date, the indicator of pastness or what have you, say absolutely nothing about the past, nor do they bear witness to it. If I am able to recollect the appointment with Paul, this episode will appear to me in the past not because of the date but because of the meaning I attribute to that date, namely the meaning of the past that I attribute to that episode. In short, for however many tags, indicators, or whatever else you may choose to attach to the memory trace, it will never contain the past.
The order of succession

On the other hand, it would be pointless to attribute the order of succession to the contents of the trace, that is something that specifies that episode B took place before episode C and after episode A. The order of succession, the very idea of a before and of an after precedes the episode, it is not part of it. If I say that B comes after A and before C, I am temporalizing A, B, and C according to an order of succession that does not depend on the elements themselves, but on the temporalizing act which I perform in accordance with an idea of succession which I already possess and which precedes the elements themselves. The relationship of succession between A, B, and consciousness is not an external relationship between atemporal elements, but rather an internal relationship between elements that are temporalized thanks to their relationship. If I say ‘I like A and I dislike B’, the relationship between A and B is external because there is nothing in A that tells me that I dislike B and nothing in B that tells me that I like A. By contrast, when I say that A comes before B, the relationship between A and B is internal because the anteriority of A presupposes in the nature of A an ‘incompleteness’ of A (instant or state) that points toward B. If A is anterior to B, it is in B that A can receive this determination. Otherwise B isolated in its instant could never confer to A, isolated in its instant, any particular quality. In the same way, in order to be posterior to A, B needs to be in some way referred to A to have the characteristic of posteriority.

Now, the question is, who is the author of this temporalization of A, B, and C? When I say ‘my last dinner with Julie was before I had a cold and after I gave a lecture in Lyon’, where does this succession come from? Certainly not from the elements themselves. My dinner with Julie, my cold, and the lecture in Lyon are just episodes without any intrinsic notion of succession. If in the memory trace of the episode ‘having had dinner with Julie’ there were ‘something’ specifying that it was before my cold and after my lecture, then every time I thought of that dinner, an order of succession would rise in my mind, which is clearly not the case. In fact I can think of that dinner with Julie without having the phenomenal experience that the dinner was before my cold and after my lecture in Lyon. If I can see an order of succession in these three episodes it is because I am the author of the temporalizing act that puts these three episodes in an order of succession. And when I say ‘I’, I mean consciousness. It is consciousness that places these episodes in an order of succession, such order is not intrinsic to the memory trace.

What conclusion can be drawn from this discussion? First of all that current theories about memory imply an insoluble contradiction which we described under the name of the paradox of the memory trace. However many experimental data may be collected, these theories will never be able to explain recollection
as consciousness of the past because from the start they break off any connection between consciousness and past. Once consciousness of the past has been separated from the past itself, the various theoretical attempts to reunite the past and consciousness are absolutely pointless. However, the memory-trace paradox does not represent the only problem with current theories of remembering. Another problem which deeply undermines the heuristic value of these theories is what we will refer to as the fallacy of the *homunculus*.

**Memory, consciousness and monitoring processes**

The assumption that the past is preserved in a memory trace contains, as we have seen, a paradox in that the past is seen to derive from present elements, but how this happens is not explained. The past is thus assumed but not explained, and any possibility of understanding the nature of recollection, true or distorted, is therefore lost. But the paradox of the memory trace is not the only problem with current theories of memory. According to some theories of memory (Baddeley and Wilson, 1986; Moscovitch, 1989; Johnson, 1991; Burgess and Shallice, 1996; Conway and Tacchi, 1996; Moscovitch and Melo, 1997), in order for the recollection of a memory or of knowledge to be correct, that is for it to be the recollection of a memory or of information that we wanted to evoke, certain selection and verification mechanisms of the memory trace must be called into play. Depending on what the subject wants to remember, these mechanisms first make a selection from the various traces stored in the memory systems, and then check whether the result of the selection meets the conditions set by the recollection task. If these conditions are not met, due to contradictions between the selection result and verification criteria, the mechanisms in question continue to make new selections until a satisfactory choice is made, a choice which does not contradict the verification criteria and which also meets the demands of the recollection task. If, for example, I remember eating in a restaurant yesterday evening, according to these theories this is due to the fact that the trace ‘dinner in restaurant’ has been selected from a host of other traces for example, ‘dinner at home’, ‘dinner at a friend’s’, in accordance with the criteria and mechanisms which guarantee the selection of the correct memory. The most widespread explanation of confabulatory symptoms provides a good example of an interpretative application of this hypothesis.

According to this hypothesis, confabulation is the result of the dysfunction of memory-monitoring mechanisms, that is, of the selection and verification mechanisms of the memory trace. The breakdown of these mechanisms is thought to prevent the inhibition of inappropriate answers, which are produced as confabulatory answers. If, for example, I want to remember what
I did last night, according to the hypothesis in question, memory-monitoring mechanisms start searching among my memory traces and make a selection, the result of which is, let’s suppose, ‘I had dinner at home’. This is then checked in accordance with criteria of plausibility and coherence with other associated memories, for instance ‘last night I went out’, and is then rejected since it does not correspond to these criteria. At this point the same mechanisms begin a new selection and check other possible memories until they find the appropriate memory, ‘last night I had dinner in a restaurant’. If, however, the memory-monitoring mechanisms are dysfunctional, ‘I had dinner at home’ is accepted as appropriate and is produced as a confabulatory memory. So according to this reasoning, the possibility of recalling an appropriate episode or meaning requires the preliminary examination of various possible answers, followed by the inhibition of inappropriate answers. Now, it is crucial to know whether monitoring processes act on a conscious, voluntary basis or, alternatively, if it is possible to assume that they operate outside consciousness and are inaccessible to it.

Are monitoring processes conscious?

If monitoring processes were conscious and voluntary, then their action would be transparent to reflexive analysis since they would be operations of consciousness. The subjective experience of reflection would, then, show me that when I am engaged in a recollection task, I consciously carry out the operations of selection and verification attributed to monitoring mechanisms. In practice this would mean that while recalling I consciously select a memory, judge whether its contents meet the conditions of the recollection task that I have set myself, reject it if they don’t and begin a new selection or accept it as a memory if it meets the conditions of the task. Reflexive analysis, however, demonstrates that this is not at all the case. If, for example, I try to recall what I did last night, I don’t need to subsequently recall a number of different memories and monitor each of them until I find the good memory, the memory that reflects what I actually did last night. Last night I watched a Woody Allen’s movie on television. I remember this without any searching and monitoring. Sometimes, however, the act of remembering needs searching. If, for example, I try to recall what I did last Wednesday at five o’clock in the afternoon, I need to search by successive approximation. I will remember, for instance, that on Friday morning I gave a talk in such and such place, that on Thursday afternoon I took the train to get to the city where I was going to give the talk, and that on Wednesday afternoon at five o’clock I was working at the talk for Friday. In other words, I operate a reconnaissance in my past, I move from memory to memory until I select one, the one which seems to best answer the question ‘what was I doing last Wednesday at five o’clock in the afternoon?’
But each of these recollections immediately appears to my consciousness as real, that is as a past image of a certain subject in which I recognize myself and to which I am intimately tied by an ontological relationship which does not allow for doubt: I was that person who was preparing the talk, taking the train, giving the talk, etc. Where I may hesitate is not on the content of my recollections, but on the date: I may not be sure that it really was Wednesday or that it was five o’clock not six. The process of selection and verification of the memory that best answers the question I ask myself does not concern the veracity of the memory, which is already given and, one could say, emanates from the memory itself; it concerns the possibility of placing it correctly in the sequence of recollections which constitutes my past. Sometimes, very rarely, I may hesitate about the source of a memory. I remember an episode, but I don’t know whether that episode was actually experienced, imagined, or dreamed. But the point is that I don’t hesitate about the fact that I am remembering an episode, but about its origin – direct experience, imagination, dreams, etc.

But what reflexive analysis shows me above all is that during recollection I never come up with memories like those produced by patients who confabulate. For example, I never come up with memories like confabulations produced by patient SD. When asked what he had done the previous day would claim to have won a running race for which he was awarded a piece of meat which was placed on his right knee. When asked to give a definition of the word ‘synagogue’ he would reply that it was something to do with physiotherapy. Now according to the hypothesis, which sees memory and knowledge as the result of the verification and inhibition of inappropriate answers by special monitoring mechanisms, in order to be able to attribute the correct meaning to the word ‘synagogue’ one would have to choose from different possible alternatives such as, let’s say ‘a church’, ‘a Jewish place of worship’, ‘a type of fruit’ or ‘something to do with physiotherapy’. Only if one manages to inhibit all the inappropriate responses will ‘Jewish place of worship’ be selected otherwise, as in the case of our patient, ‘something to do with physiotherapy’ or any other answer could be given. This reasoning also implies that if, before his amnesia, the patient was asked what he had done the previous day, he would have consciously examined and rejected the possibility of having won a running race and being awarded a piece of meat which had been placed on his right knee and would instead have produced the correct answer. There is no evidence that, in order to recall an episode or a piece of information, subjects are engaged in consciously evaluate and reject candidate memories and information of the type described above.

Certainly confabulation is more often far more plausible than the examples given of the running race and the definition of ‘synagogue’. In fact, most patients who confabulate produce confabulatory memories which cannot be
distinguished from ‘real’ memories by an interlocutor who does not know the patient’s history and current situation. But even in the case of the most plausible confabulation, like the one produced by patient MG (see above), it is difficult to imagine that on asking someone in hospital why he is there, this person, when answering, should consciously consider, among others, the possibility of being in hospital to accompany a friend. But even if that were the case, that is if recalling an episode or a meaning were possible due to conscious and voluntary inhibition of inappropriate answers, something which reflexive analysis leads us to exclude, it remains to be decided on what basis the correct answer is selected. One criterion could be that of plausibility. Less plausible answers are inhibited while the more plausible are accepted. But isn’t plausibility itself a meaning which is attributed to a possible answer, in which case shouldn’t plausibility also be subject to a process of verification like any other meaning? And shouldn’t the criterion which I use to determine plausibility also be verified, and so on endlessly, like in a game of Chinese boxes? It is not worth arguing (Johnson and Raye, 1981; Johnson, 1988; Johnson, 1991) that in recollection the correct choice from various alternatives is based on the evaluation of the qualitative characteristics of information, for example the amount of perceptual detail and the quantity and type of memories associated with the episode we want to recall. Indeed, is it once again a question of quis custodet custodes? Who can assure me that when I remember meeting Paul the other evening, the bright red of his pullover is a perceptual detail which comes from my memory and not my imagination? Why should I grant these so-called perceptual details the status of veracity which I deny the memory itself? Similarly, who can assure me, for example, that the fact that in my memory Mary is with Paul is an ‘associated memory’ which guarantees the veracity of the main one, that is the meeting with Paul, and not a confabulation? As you can see, there is no way of escaping this circularity.

**Are monitoring processes unconscious?**

The subjective experience of reflection thus leads us to exclude the notion that, with some rare exceptions, the processes of information selection and verification are voluntary and conscious. Are they then processes which work outside consciousness and which are inaccessible to it? But to accept the hypothesis of unconscious monitoring mechanisms means falling into what we call the homunculus fallacy, that is the contradiction of postulating the existence of a sort of unconscious consciousness, that is of unconscious monitoring mechanisms endowed with intentionality which select, evaluate, and reject false memories and provide conscious consciousness with only real memories. It is unclear on the basis of what theoretic assumption one can attribute intentionality to
this kind of process. Since Brentano (1874 (1973)), intentionality has been a characteristic of consciousness and, as Husserl (1950) says, represents the need of consciousness to exist as consciousness of something. Attributing intentionality to an unconscious process, then, is equivalent to giving unconsciousness the attribute of being subject, that is attributing consciousness to the unconscious. Moreover, this unconscious consciousness is inaccessible to the real sense of consciousness. As in a game of Chinese boxes, one personality would contain another personality, a homunculus to be precise, endowed with a shady, inaccessible consciousness which is busy resolving problems like rejecting false memories and is ready to provide conscious consciousness with the result of this detailed selection. But let’s examine this problem more closely and see why it is illusory to attribute intentionality to the unconscious.

**The fallacy of the homunculus**

Let us first of all clarify what is meant by intentionality. I am conscious of this glass on the table. In other words, I have perceptual consciousness of glass-on-the-table. I also perceive the glass from a certain angle, not only visual but, let’s say, existential: it exists before me, it presents itself to me in a certain perspective and, as an object among other objects, it has a certain relationship with these. Here again, this relationship is not only spatial (proximity, distance, adherence, etc.) but also functional, aesthetic, etc. In short, the being of the glass-on-the-table is a certain, specific, and original being which my consciousness is at this moment selecting.

What can be deduced from this? If we were idealists, we would say that the glass-on-the-table exists as ‘content’ of my consciousness, which in some way absorbs it. If we were realists, we would instead tend to consider the glass-on-the-table as an element of a reality which precedes and transcends consciousness, and to which consciousness can only adhere. And, in his cognitivist diversity, the realist would say that this information, ‘glass-on-the-table’, entered my cognitive system where it underwent a series of computational transformations before becoming perceptual consciousness of the glass-on-the-table. Nevertheless, if we free ourselves from what Sartre (1939) called alimentary philosophies, that is, from realism’s tendency to make the object ‘eat’ the subject and idealism’s tendency to make the subject ‘eat’ the object, we discover the true relationship between consciousness and world, between subject and object. These are born together. They are involved in a relationship where, although they retain their autonomy, neither term can exist without the other. The result of this relationship between subject and object, between consciousness and world, is intentionality, that is, the need of consciousness to always be consciousness of something.
Intentionality then, as relationship between consciousness and world, is a basic characteristic of consciousness which has no ‘inside’ but which defines itself by being ‘outside’, intentionally oriented toward things. At this point we must ask whether intentionality can also be attributed by right to the unconscious. Searle (1983), while rejecting functionalist positions, draws a distinction between consciousness and intentionality, maintaining that there can be non-intentional conscious states, for instance a sudden feeling of excitement, and unconscious intentional states, for example those beliefs which I am not thinking of at this moment but which determine my behavior. But why should a sudden feeling of excitement be a state of non-intentional consciousness? Excitement, or its opposite, depression, are states of consciousness, as Searle calls them, on the same level as any other consciousness. Like any other state of consciousness, that of excitement or depression does not escape its own law, that is of being destined to come out of itself in order to come to terms with the thing, that world which it is trying to reach and which continues to escape it. Accepting the existence of non-intentional states of consciousness is the same as making consciousness coincide with the thing. Non-intentional consciousness is quite unthinkable because it would be consciousness deprived of its original essence, that is, the negation of itself as ‘thing’. One must not fall into the trap of thinking that first I am excited and then I am conscious of the glass-on-the-table, or that a sudden state of excitement joins my perception of the glass-on-the-table from outside. What one should say is that in a state of excitement I am aware of the glass-on-the-table. Mine is excited consciousness of the glass-on-the-table. Therefore, once again there are no exceptions to the law of consciousness: all consciousness is intentional in that it must, by nature, be consciousness of something.

Searle maintains that the idea that consciousness must exist as consciousness of something neglects a fundamental distinction, ‘when I consciously experience anxiety, there is something which my experience is experience of, that is anxiety, but this sense of “of” is radically different from the “of” of Intentionality which appears, for instance, when I affirm that I am consciously afraid of snakes; since, in the case of anxiety, the experience of anxiety and the anxiety itself are identical, while fear of snakes is not identical to snakes.’ But in proposing this example Searle confuses the essential structure of a reflexive act with that of a non-reflexive act. He seems to ignore the fact that there are always two possible forms of existence for consciousness: non-reflexive consciousness, that is which appears when ‘I am afraid of snakes’ and reflexive consciousness, that is consciousness of me-as-anxious. These are two different types of consciousness, but both are intentional. The trick Searle uses involves comparing two different degrees of consciousness. In the case of
anxiety, he makes the example of reflexive or second-degree consciousness; I am aware of being anxious. But first degree or non-reflexive consciousness is only anxiety, anxiety which opens onto the world like any consciousness and like any consciousness it is intentional in that it is outside itself, as anxiety, toward the world. It is by no means true that 'in the case of anxiety, the experience of anxiety and anxiety itself are identical'. The anxiety which pervades me in this case is the object of my consciousness, which makes itself conscious of anxious consciousness. Without this second-degree reflexive act, which gives me a point of view on my consciousness, my anxiety is my consciousness without my being aware of it: I will anxiously drink from the glass-on-the-table without being able to avoid the intentionality of my consciousness which will be anxious consciousness of the glass-on-the-table. There is, then, no difference regarding the degree of intentionality between my consciousness of fear of snakes and my consciousness of anxiety. They are both intentional. The difference is that the former is non-reflexive consciousness while the second is reflexive consciousness, that is consciousness which makes anxious consciousness its intentional object.

If we then analyze the possibility maintained by Searle of unconscious intentional states, that is the so-called beliefs to which we fall victim without being conscious of it, it is clear that these supposed beliefs are not at all that unconscious background which Searle talks about, but they constitute consciousness itself. Searle says 'I believe that my paternal grandfather never once left the American continent, but until now, I had never formulated or clearly examined this belief'. What was this belief of Searle’s before it became conscious? Searle maintains that it already existed in his unconscious in its intentional form of ‘my grandfather never left the American continent’. This seems to be a rather rushed interpretation of the facts. Essentially Searle is saying that, in some way, the fact that his grandfather never left the American continent is something that he had always known, even if this thought has only now become conscious. If, then, it has always existed it must have been stored somewhere, and where or what is this place where thoughts are stored before becoming conscious, if not the unconscious? There is almost no need to emphasize the absurdity of such a hypothesis. Even if the unconscious does exist, who will believe that it contains pre-established thoughts and beliefs? This type of explanation, which has a long tradition in psychology, stems from the fact that spontaneity of consciousness is not accepted and thus is traced back to an ‘elsewhere’, the unconscious, the cognitive system, etc., where already formulated thoughts are believed to exist in a passive mode, in propositional form, which will then become conscious. What has not been seen is that such a hypothesis, in trying to trace the spontaneity of consciousness to another,
unconscious location, only puts off the problem of the existence of beliefs and thought in general, a problem which sooner or later will have to be formulated. An unconscious and intentional thought or belief does not only present the paradox of passivity, that is of the existence of inactive intentional states, but it also implies another question: who, in this unconscious, believes, has believed, and will continue to believe that Searle’s grandfather never left the American continent? The problem of the existence of an unconscious subject who believes thus arises. We shall come back to this problem shortly. For the moment it is enough to emphasize how even if an unconscious subject were admissible, he or she would in turn have to draw his or her thoughts and beliefs from another ‘elsewhere’, because if the spontaneity of consciousness is not accepted, there is no reason to accept spontaneity of the unconscious. The attempt to trace the origins of spontaneity to an unconscious elsewhere is nothing but an absurd, naive attempt to free oneself from the anguish we feel assisting the unceasing ex nihilo creation of consciousness which we do not create ourselves. You could object that consciousness of ‘grandfather-never-left-the-American-continent’, however spontaneous we may consider it, is still coherent with a given fact, that is, that Searle’s grandfather never actually left the American continent. However, we will have to ask ourselves where this consistency between consciousness and reality comes from. There is no need to resort to explanatory idols such as the existence of unconscious beliefs. Consciousness, as intentional consciousness of the world, can only be a perpetual synthesis of past and present consciousness. In short, it is not by resorting to mechanistic determinism, which claims that consciousness comes from unconscious intentional states, that we will be able to understand spontaneous, intentional existence. Saying that consciousness is the synthesis of states of past consciousness with present consciousness does not mean tracing its origins to somewhere else, it simply means acknowledging to consciousness an existence ex novo in accordance with a certain synthetic form.

As an example of unconscious intentional states, that is of background, Searle gives that of Carter who formulates the intention of presenting himself as a candidate in the presidential elections of the United States. To be more precise, Searle maintains that ‘an intentional state does not determine its conditions of satisfaction and thus is not the state which is intertwined in a Network of other Intentional states and on the Background of pre-intentional practices and suppositions which in themselves are neither Intentional states nor parts of conditions of satisfaction of intentional states’ (Searle, 1983). In order to express his intention of presenting himself at the presidential elections, Carter has to know that the United States is a Republic, that there are
regular presidential elections where two large parties are rivals in the race for presidency, etc. Whoever presents himself as a candidate has to have the backing of his/her party, be actively supported by a large number of people, and finally, in order to become president must be voted for by the majority of voters. Well, concludes Searle, even if none of these states has a direct link with Carter’s intention to candidate himself, he could not form such a plan, that is ‘the intention of presenting himself for the American presidential elections’, without such a network of intentional unconscious states. ‘One could say that his intention makes reference to certain other Intentional states, inasmuch as it can only reach the conditions of satisfaction it does, and consequently only be the intention it is, because it is located in a Network which includes other beliefs and other desires. Besides, in every real life situation, beliefs and desires are only part of an even greater complex of other psychological states: there are subsidiary intentions such as hope, fear, anxiety and expectation, feelings of frustration and satisfaction. In short, what I call ‘Network’ is nothing other than this holistic network’ (Searle, 1983). What Searle wants to explain is that our acts are not fortuitous but presuppose a ‘Network’ of meanings which condition their realization. ‘We can easily understand what someone’s intention to become president is, but we cannot by any means understand the intention of becoming a cup of coffee or a mountain, because – among other reasons – we do not see how such an intention can be consistent with the Network’. What is difficult to understand is how this network can exist and how it can be made up of unconscious intentional states. In fact, if we accept, its existence we also have to face the problem of its origins. What is the nature of this unconscious and intentional Network? Even though he vigorously rejects every functionalist hypothesis, Searle seems to make a similar mistake to that which he accuses the functionalists of making. Like them, he resorts to the unconscious in order to explain consciousness. Searle’s is not an unconscious made of information processing, but one of beliefs or meanings which condition conscious life as much as the functionalist unconscious does. But Searle’s hypothesis is presented with exactly the same problem as is functionalism: that of explaining, once and for all, how man’s conscious life can be the result of something which happens in the unconscious in accordance with rules fixed by consciousness, Searle’s in this case. We shall come back to the illusiveness of considering consciousness as an ‘effect’. For now, it is enough to understand that unless its ontological origin is clarified, Searle’s Network is nothing but a figure of speech. Actually, there is no apparent reason to make reference to an unconscious intentional Network of meanings to explain the fact that formulating the intention of presenting oneself for presidential elections is plausible and
comprehensible, while formulating the intention of becoming a piece of amethyst is not. As we have already said, consciousness cannot but be a perpetual synthesis of earlier states of consciousness.

My present consciousness is new consciousness which is born ex nihilo and does not resemble anything before it. But this does not mean that my earlier states of consciousness are not synthetically represented in it. Soon I will leave the house to go to the cinema. This will happen, not because I have a sort of redeeming Network which allows me to do this but forbids me to leave the house to hunt a jaguar, but simply because my present consciousness is mundane consciousness of a being which is me, a being which carries around its past without any possibility of getting rid of it. There is, then, no Network of beliefs and meanings on one side and a consciousness which is determined by it on the other side. My consciousness is the network of the beliefs and meanings of which it is made up. Beliefs and meanings which are not in an unconscious elsewhere in a strange state of active passivity, that is in an oxymoron – meanings and beliefs whose origin as well as the present state need explaining. In fact either they are passive in the unconscious and so have no way of influencing conscious life or they are active, in which case, where their activity is supposed to stem from needs explaining. Carter’s consciousness in formulating his intention of presenting himself at the presidential elections is such as it is, also because at a certain point in his life Carter became conscious of living in a country where there are presidents, presidential elections, etc. His present intention is not the result of unconscious intentionality made up of unconscious beliefs, but is the present form of his present consciousness which implies, and is, the synthesis of all his earlier states of consciousness.

If I now take a cigarette to my lips and inhale the smoke, this does not happen as the result of a series of unconscious presuppositions. It is not because I have the unconscious and intentional belief that, for instance, ‘there are objects called cigarettes of a certain shape which serve a certain purpose, which are bad for your health but also give pleasure, which in order to smoke you must go through a series of movements which involve the contraction of certain arm muscles but not of others, and that each movement consumes a certain amount of energy which is reflected in the consumption of oxygen and carbon compounds, etc.’ This is only one of the countless explanations which I can give for my act of smoking, or to put it another way, one of the possible ways my consciousness thematizes my act of smoking in order to assign it a meaning. Meanings and beliefs do not precede consciousness, if anything they follow it, that is they are possible thematizations of the world on behalf of consciousness. Consciousness itself has no need for meanings and beliefs which precede it and guide it out of the darkness of the unconscious. Consciousness is the
synthetic combination of these meanings which, with a secondary operation, can become theses for consciousness which, in this case, ceases to be them in order to detach itself from them and make them object of its own thematization. In short, there may be a metaphysical problem in consciousness but there is definitely not an ontological problem: my present consciousness is made up of my earlier states of consciousness. We will see later, when our discussion leads us to the problem of temporality, what exactly is meant by this. For now, it is enough to have clarified the impossibility of supposing the existence of an unconscious intentional world.

We have just seen that hypothesizing the existence of intentional, unconscious beliefs involves the problem of supposing the existence of an unconscious subject who ‘believes’ something (for example that Searle’s grandfather never left the American continent). This unconscious subject, then, is supposed to reveal itself just when the conscious subject is ready to thematize this type of pre-packaged thought, stored in the frozen unconscious, together with the unconscious subject, passively waiting for consciousness to defrost it, as if by magic. This type of interpretation, which does not seem to deserve comment, is however worth analyzing once and for all.

The first question to be asked clearly concerns the nature of the subject, this unconscious who believes. A first objection could be raised at this point, and that is that in reality questioning the nature of an unconscious subject is the same as posing a false problem. In fact, there is no need to presuppose an unconscious subjectivity as author of acts like believing or thinking. Beliefs and thoughts could be unconscious without needing to be the beliefs and thoughts of someone, that is of an unconscious subject. They would be kinds of pure representations which in themselves were complete and self-sufficient. In other words, that network of meanings which, according to Searle, ‘are in one’s head’.

Let us suppose that shortly I leave the office to go and buy cigarettes at the bar on the corner. According to the hypothesis in question, this project of mine is plausible inasmuch as a series of unconscious beliefs make it so. The belief, for instance, that there are shops that sell cigarettes, that these shops have certain opening hours compatible with the time I go to one of these to fulfill my objective, that I am a great consumer of cigarettes, and so on. Good. These beliefs, which make my objective possible, supposedly lie passively in the unconscious, waiting to intervene, and guide my behavior when necessary. Question: who informs them when it is their turn? How do they come out of their passive state to become active? You might say they are activated, as functionalist literature does. A group of representations or of neurons are activated in order to attain said goal. But, once again, who activates these representations?
Certainly not my consciousness. The moment I go to buy my cigarettes my consciousness is pure, non-reflexive consciousness, consciousness without a subject, consciousness of-cigarettes-to-be-bought. They could be said to activate themselves, that is, to pass from their passive state to an active one without the intervention of any external factor. But why them in particular? Why should only those beliefs which are useful to one’s purpose be activated, and not others? You could object further that it is precisely the need to reach one’s goal which activates them. But even if we accept this kind of reverse teleology, which in any case needs clarifying, who is it that goes and tells the beliefs necessary for the purpose that it is they, and not others, that are necessary? There could be said to be a system of beliefs which is in some way always active in the unconscious and ready to face the needs of the moment. But how could a system of beliefs which is in a state of continuous activation, in a sort of erethism work? In a similar condition, there would be beliefs in contradiction with each other and in an equivalent state of activation. For example, the beliefs ‘there are shops that sell cigarettes’ and ‘man has been on the moon’ would be activated contemporarily. If both these beliefs are activated at the same time, who can assure me that halfway from the office to the bar I will not decide to go to the moon? This demonstrates, then, that a problem of priority among beliefs arises, dictated by the necessity of the moment. And this priority takes us back to the problem of selecting the appropriate beliefs; since this is not carried out by a conscious ‘who’, the existence of an unconscious ‘who’, author of this choice, must be presupposed.

So here we are, once again, struggling with the nature of a who, an unconscious subject who intentionally turns to an unconscious world made up of beliefs, meanings, thoughts, memories, and so on. Who is this unconscious subject, this homunculus that lives in the shadow of the unconscious, master and origin of conscious life?

The problem of an unconscious ‘who’

First of all it is plain to see that the unconscious homunculus is distinct from the object it intentionally addresses, beliefs, memories, etc. In a certain sense then, we find the same relationship between the homunculus and the object it addresses as that which exists between consciousness and its object. In short, one could say that the homunculus is conscious of beliefs, of the meanings of the thoughts and memories it has to select, activate, etc. But what is the unconscious homunculus conscious of? What is the nature of unconscious meanings, memories, beliefs? Let us go back to the case of memory, which is that which interests us most. In its work of selecting and checking memories, the homunculus deals with both true and false memories. Its task is that of selecting memories,
sending the real ones to consciousness and rejecting the false ones. Now, it is obvious that if we accept such a hypothesis, we must also accept that the memories that the homunculus examines or, in functionalist terms, the information that the monitoring systems elaborate, must be already specified in syntactic-grammatical terms. Since no active role is attributed to consciousness, all that arrives there must already exist first, in the unconscious, in the form that it will have afterwards in consciousness. For example, our patient’s confabulatory memory of ‘having won a running race’ will be exactly the same in the unconscious, in a sort of state of inactivity, waiting to be evaluated by the homunculus as a possible memory. In this way the homunculus fallacy not only postulates the existence of an unconscious consciousness, that is, it is based on an oxymoron, but since this unconscious consciousness must have an object to turn to, it also postulates unconscious mnemonic activity already specified in syntactic-grammatical terms. In other words, as in a photocopy of conscious life, memories are found in exactly the same form in the unconscious, ready to be selected by our homunculus, and sent to that kind of passive container which is thought to be consciousness.

At this point one may object that memories do not need to exist in the unconscious, in the form that they have in consciousness, because consciousness can be considered as having emergent properties. Searle (1983), for example, maintains that the liquidity of water, just like consciousness, digestion, photosynthesis or cell reproduction, is caused by elements which precede it in the causal chain and takes form in the structure which regulates the relationship among these elements. But the molecular structure of water does not, by any means, reveal its liquidity. If I look for its chemical structure, what I find is a certain relationship between pieces of matter which are called atoms, and I express this relationship with a conventional symbol: H2O. In H2O there is no liquidity, just as there is no solidity or gaseous state. In order for H2O to be liquid, something which has nothing whatsoever to do with its molecular structure is required. That is, H2O needs to be found at a certain temperature and under certain pressure conditions. But even if the conditions of temperature and pressure are met, we still have not discovered the liquidity of water. In order for water to be liquid, consciousness must notice it. Without consciousness bothering to discover water as liquid, water, assuming its chemical structure is known, is nothing but H2O. If its chemical structure is unknown, water is simply water, that thing we are all familiar with and which we know has certain characteristics, among these liquidity, a term we use to distinguish one state of matter from another. As for temperature and pressure, they only become significant when a consciousness which knows a bit of physics recognizes that these two variables are essential in order for water to be liquid.
In other words, the molecular structure of water is neither necessary nor sufficient for its liquidity. It is not necessary as water can be considered to be liquid before, and quite apart from, any knowledge of its molecular structure. It is not sufficient since even if by some absurdity there were someone in the universe who knew the molecular structure of water before knowing that it was liquid, from the former he could never derive the latter. Searle’s causality could, in reality, be easily turned upside down. It could be maintained that it is the molecular structure of water that is caused by, and realized in its liquidity. It is because water is liquid that it has a certain molecular structure. But if we decide, once and for all, to leave aside the circularity of causal thought, it is clear that neither the liquidity of water nor its structure can lay claim to priority over the other. They both come into being as meanings, that is, as objects which present themselves to consciousness in accordance with different points of view. Water is both liquid and H2O without one of these being cause of the other.

In short, in order to maintain that liquidity is an emergent property of H2O and that conscious remembering is an emergent property of memories that exist in the unconscious in a form that is different from the form they have in consciousness, one should show how these properties, liquidity and conscious remembering, emerge from different states. In the absence of a clear specification about the passage from one state to another, the liquidity of water and conscious remembering are nothing but a figure of speech.

Turning back to the problem of the homunculus, we see have gained nothing in supposing its existence. In fact, we find ourselves with an unconscious made up of the same elements we left in conscious life: on the one hand a subject with its baggage of intentionality, the homunculus, on the other a pre-fabricated unconscious world that the homunculus intentionally addresses. At this point it is clear that what we had invoked to explain conscious memory requires an explanation itself. In fact, if the existence of unconscious monitoring mechanisms, the homunculus, has been set as a condition to explain the conscious emergence of memory, the criteria in accordance with which these unconscious monitoring mechanisms are supposed to operate also need explaining.

In other words, on the basis of what criteria does the homunculus distinguish a true memory from a false one? In order to reject false memories, the homunculus needs to know what it is rejecting. The homunculus in fact has to choose among memories and in order to do so it must be able to represent these memories to itself. Otherwise how could it accept true memories and reject false ones? But it is not sufficient for the homunculus to be able to distinguish between memories. It must also recognize false memories as ‘to be rejected’ and that implies that the homunculus must be able to represent its own activity to itself. Indeed, how could the homunculus recognize memories ‘to be rejected’
without being conscious of recognizing them as to be rejected? In short, the *homunculus* implies not only a conscious representation of memories ‘to be rejected’, but also a consciousness of what the *homunculus* must do in order to carry out its action correctly. But what kind of self-consciousness is the self-consciousness of the *homunculus*? It must be consciousness of the tendency to reject false memories, but precisely for the aim of not being conscious of that tendency. An unconscious consciousness that is conscious of itself in order not to be so. At this point wouldn’t this *homunculus* in bad faith (Sartre, 1943) also need its own unconscious *homunculus* to help it distinguish real memories from false ones? The game of Chinese boxes continues and what we are left with is only a *reductio ad infinitum* which is clearly of no explanatory value.

**Is a scientific theory of consciousness possible?**

Up until now our discussion dealt with present theories of memory and their possible application to the interpretation of confabulation. We have shown that these theories are based on a paradox, that is, on the idea that the past is passively preserved in memory traces, and on a fallacy, the fallacy of the *homunculus*, that is, the assumption of the existence of a sort of unconscious consciousness in bad faith. So, memory traces and monitoring processes are just explanatory idols of which theories of memory should get rid. The problem is that there is a tendency to think of memories as unconscious items that one brings to consciousness. But we have to understand consciousness before we can assume that memories simply ‘rise’ to it or are tacked onto it. In particular, it must be ascertained whether memory and consciousness are part of the same structure or not. The question we have to ask at this point is whether a scientific theory on the relationship between memory and consciousness is possible.

The aim of science is to achieve objective knowledge and its method the establishment of quantitative relationships, the consequence is that a scientific theory of consciousness will have as aim the objectivation of consciousness and the determination of the quantitative relationships that describe it. In other words, in order for a science of consciousness to be possible, consciousness must be a measurable object which is independent of the observer and of the point of observation. Yet, to objectivise consciousness, i.e. to consider it as an object, means to transform consciousness into something it is not. Consciousness by definition is subjectivity that describes itself according to rules of quality and not of quantity. This flowerpot standing in front of me, the memory of meeting Paul last night, the blue I am now imagining are present to my consciousness, establishing that subjectivity that is me. And I am the only being responsible for this subjectivity. Of course, at the same time that being
that I am subjectively is not only for itself but also for the others. I am continuously observed, object for other consciousnesses like other consciousnesses are objects for myself. However, this objective consciousness has not the characteristics of consciousness because it is an object and my consciousness is not object but subject. The only being that for me has the characteristics of consciousness is my consciousness. And even when my consciousness becomes object not for others but for me, i.e. when with a reflective act I become conscious of my consciousness, the consciousness I am conscious of is a me which is not me, that does not have the characteristics of consciousness. The objectivation of consciousness, therefore, reflects a radical metamorphosis. Even if I could see myself distinctively as an object, what I would see would not be an adequate representation of what I am, but only the perception of the distance that separates me from myself, that is the objective perception of my being other, which is radically different from my being for myself, that is from that being that I am in the irreflective mode. My considering myself honest, for example, does not describe at all what I am for myself, because I cannot be honest for myself, that is, in the irreflective mode of consciousness. My considering myself honest, for example, does not describe at all what I am for myself, because I cannot be honest for myself, that is, in the irreflective mode of consciousness. The qualification of 'honest' characterizes me as object: it is that being of which my consciousness is conscious that is honest, not what I am in the irreflective mode, that is my consciousness.

Even considering my consciousness as object for the consciousness of others, we can see that this does not establish how I am for myself but how I am for others. When somebody describes my character saying that I am 'jealous', 'nice', or 'irascible', I don’t recognize myself at all in this description, nevertheless I know that it is me, an objectual 'I' that I cannot deny being but that does not describe at all what I am for myself. In this way the 'I-object', or objective consciousness, establishes a knowledge that does not refer to the reality of consciousness, that is, to what I concretely am as subject, but to a metamorphosed consciousness, a consciousness transformed into object. And this is the metamorphosis of consciousness that science requires in order to establish consciousness as the object of its investigation. Therefore, any scientific theory of consciousness can only be a theory that concerns objective consciousness, that is a degraded form of consciousness, and will have to renounce, by definition, the possibility of having ‘fresh’ information concerning consciousness because consciousness in its entirety defines itself precisely as what is not object.

Earlier in this chapter we argued that intentionality is a characteristic of consciousness since it describes the necessity of consciousness to always be consciousness of something. We would now like to add that consciousness is always consciousness of something in a certain way. This means that consciousness takes a certain point of view of its object and of the same object
consciousness can take various points of view. The first consequence of this statement is that different modes of consciousness exist and each of these represents an original and irreducible mode of addressing the world. By 'original' we mean that each mode of consciousness is different from any other. By 'irreducible' we mean that the existence of each original mode of consciousness cannot be traced back to something else that precedes it in causal or ontological terms. A taxonomy of the different modes of consciousness is far beyond the aims of this work. Our aim is rather to operate certain distinctions that will allow us to see, though at a rather general level of description, the differences between two modes of consciousness, namely what we will indicate as knowing consciousness (KC) and temporal consciousness (TC).

**Knowing consciousness and temporal consciousness**

KC describes what is usually referred to as semantic memory. A substantial difference, however, distinguishes KC from what is presupposed in the idea of semantic memory. Unlike semantic memory, KC is not based on the idea of unconscious mental representation. KC is consciousness of the object, it is already outside itself and directed toward the object to be known. And it is consciousness of the object in a certain mode, an original mode that makes it impossible that what I know, for example, can be mistaken for what I remember or imagine. It can be argued that I could never, for example, recognize this packet of cigarettes as red if I had not already somewhere in my mind and my brain an idea of red. But this kind of argument is unintelligible because it presupposes the existence and the activation of unconscious representation and in so doing falls into the fallacy of the homunculus. If I recognize the packet of cigarettes as red it is not because I operate, unconsciously, a sort of correspondence between the packet I perceive and a mental representation that I carry in me. What one should say is that I am conscious of this packet of cigarettes as red because my present consciousness, since it is a synthesis of all my past consciousnesses, is also consciousness of red. But I can also see this packet of cigarettes as ugly, dangerous, attractive, almost finished. In short, the meanings I attribute to the packet of cigarettes are concretely infinite and depend on my being conscious here and now as synthesis of what I have been before and elsewhere. In this sense KC is temporal since it is a synthesis of what I have been and we are our past or was gewesen ist, as in Hegel. But at the same time KC, which is temporal, is also atemporal in the sense that the time of which it is made is not recognizable. KC is the past but it is not consciousness of the past, nor of the present or of the future. There is no time in the packet of cigarettes in front of me although it is thanks to time, that is, of the past I am made of, that I can see the packet as red, dangerous, etc.
Consciousness of time, i.e. the thematization of the object in the mode of temporality, is what we have called TC. TC, as far as it poses its object according to the structures of time, i.e. the past, the present, and the future, is an organized, original, and irreducible form of consciousness, for addressing the world. Unlike KC, TC transcends the mere presence of the object in order to set it in time. When I say that I remember yesterday’s dinner, that now I am in my office, and that later I will go out to buy some cigarettes, knowledge is presupposed, as we will see later, but it is not the aim of the temporalizing act that I am making. In order to remember yesterday’s dinner, to affirm that I am now in my office, and to plan the act of buying cigarettes, I have to know what a dinner, an office, and cigarettes are. However, this knowledge is only the structure on which my act is founded. It is through knowledge that the temporalizing act is realized but it is not for knowledge that it is realized. Although TC is an organized structure of consciousness because past, present, and future are not isolated dimensions but continuously refer to one another, consciousness of the past, of the present, and of the future are nevertheless subordinate structures of TC, and for the sake of clarity, we will describe them separately.

Consciousness of the past is an act through which the object of consciousness is seen as absent, or as non-present. But, in contrast with imagining consciousness, which sees the absent object as non-existent, for consciousness of the past the object is absent in the past. It is this consciousness of the past that we call remembering. But consciousness of the past is not consciousness of a generic and impersonal past. When I say that I remember that Kennedy was killed in 1963 in Dallas or that Dante wrote the Divine Comedy, I am not conscious of the past but of a piece of information, a notion, that joins other notions to form a certain type of knowledge. Kennedy is president of the United States and past, Dante is a poet, Italian and past. In other words, it is about a generic and impersonal past, not my past. In order for consciousness of the past to exist, i.e. remembering in the mode of TC, there must be a deep link between the being I am now and a being I was before. It is to that past being of mine that consciousness of the past refers, and it is that past being that remembering represents.

In order for consciousness of the past, i.e. remembering, to exist, there needs to be a present of which the past is past. It is present consciousness that is consciousness of the past. It must be emphasized here that by ‘present’ we mean subjective or phenomenological present, which distinguishes itself from ‘objective’, or clock-measured present. Objective present, as measured in neurophysiological experiments for example, lasts a few milliseconds and is practically instantaneous present. Accordingly, present consciousness would also last a few milliseconds and as a consequence it would be either instantaneous or always ‘behind’ already in the past. However, the subjective phenomenological
present cannot be measured in objective time. The phenomenological present is not an instant, but a dimension that distinguishes itself from the clock-measured present with its characteristic of presence. So present consciousness is not an instant, or something that is always already past, but is the relationship between the subject and its object in the mode of presence. But present consciousness can also be consciousness of the present. Consciousness of the present is an act that temporalizes the world in the present form and that does not confuse itself with perception which is the mere presence of consciousness to the object. There is no temporal dimension in my perception of this chair. It is only when I feel myself contemporary to this chair, to this table, to the city, to the world that the present comes to the world as consciousness of something in the present mode. So consciousness of the present, far from coinciding with perception, represents thematization of perception in the present mode.

The third subordinate structure of temporality is consciousness of the future. Consciousness of the future puts its object in the future mode. It is thematization of my probable in the future mode. By probable we mean what is founded on knowledge of my past and present, and this radically distinguishes consciousness of the future from wishes. Tomorrow I will be in Paris and I will continue working on this paper. This is one of my probable, a thematization in the future mode of my being that founds itself and is synthesis of knowledge of my being in the world: I know that I live in Paris, that I am writing this paper and that tomorrow is not Sunday. When instead I say ‘I would like to be an astronaut’ I don’t express a probable, but something that transcends the knowledge of the mundane being that I am.

So TC opens the possibility of a temporal existence for the subject. A subject who is conscious of a personal past when remembering a personal present, in which he/she is oriented, and a personal future, in which he/she is projected. But before proceeding further and see what is the relationship between TC and confabulation, which is what interests us more, we need to know what the normal operations of TC are.

### Uniqueness and multiplicity

Consciousness, we have seen, must always be consciousness of an object. One of the characteristics of the object of consciousness is that it represents and reveals to consciousness a uniqueness (U) and a multiplicity (M). Let’s consider this point more closely. This pen on the desk is both a pen and the pen. In the first case, it is an undetermined pen, something that belongs to the category of ‘pens’, an object that I recognize and use appropriately because I recognize it. In contrast, in the second case ‘the pen’ is a determined object, it is
exactly this pen in front of me, the pen I bought yesterday and that I will be using tomorrow. So the pen reveals a \( U \) and a \( M \). \( M \) is reflected in its being a pen and not a different object. \( U \) manifests itself in its being this precise pen and not another. However, the \( U \) and \( M \) of the pen are not adjacent qualities, external to one another. There is a sort of hierarchy that establishes the relationship between the object as representative of an \( M \) of objects and the object as representative of itself, in being \( this \) object that distinguishes itself from all the other objects of the world. The \( U \) of the object, of this pen, of the room, of my feeling happy is \( this \), i.e. they manifest a \( U \) that distinguishes them from the \( other \) pen, room, or feeling of happiness. But their \( U \) is founded on an \( M \), the \( M \) that acts in such a way that these objects manifest themselves to consciousness as unique but under a certain, already distinguished form, i.e. pen, room, happiness. If we consider the opposite condition, we see that for the undetermined object there is no need to represent the \( U \). I can think of a flower, a love, a city, a world without these objects needing to be \( this \) flower, \( this \) love, etc.

But what happens when in front of me there is an object I have no experience of, that I don’t know? At first sight one would say that any object that I see for the first time is unique and doesn’t carry in itself that \( M \) that we attributed to known objects. But in actual fact, as soon as an unknown object shows up, all the \( M \) of which it is made manifests itself instantaneously. This geometric form that I have never seen before is already ‘geometric form that I have never seen’, that is it summarizes and represents in the negative form what my consciousness has been before. And this can happen without my needing to notice the \( U \) of the object. So the object, be it known or unknown, is first the expression of an \( M \) and then becomes \( this \) object, i.e. declares its \( U \).

However, in order for the \( U \) of the object to be revealed, it must be addressed by consciousness in the mode of TC. In other words, in order to achieve the passage from the \( M \) to the \( U \) of an object, there must be an act of consciousness that selects its object as a \( this \), and this can happen only if consciousness transcends the knowing mode to put itself at the level of the temporalization of the object. One would certainly want to argue that in what we have called KC the object also appears to me as unique. The cup in front of me is in any case \( this \) cup, it is with this cup that I interact and not with a generic cup when I pick it up and bring it to my lips. Of course, but its \( U \) is revealed only if I consider it as a \( this \), that is when I notice its presence as an object that is not my consciousness, and to which my consciousness is contemporary. When I simply use it, when I execute the appropriate sequence of gestures in order to pick it up and bring it to my lips, or when I look at it wondering whether it is full or empty, there is no \( U \) in my consciousness. I do not even need to be conscious of the cup in order to use it and when I wonder whether it is empty or full I am aware of a problem to be solved, not of the presence of \( this \) cup. In short, there is the
continuous possibility of varying the relationship between $U$ and $M$ and this possibility depends on the mode in which consciousness addresses the object.

However, one could ask, where is the evidence that in TC the object manifests itself to consciousness as unique and not according to its $M$. I can easily, for example, remember a generic walk along the shore without being able to locate it in any specific time but only in the past in general. But this does not mean that the walk I remember is not a particular walk but instead the general idea of walking along the shore. The fact that in this memory certain details are missing, where and when, for example, does not mean that I am not remembering a specific walk. Remembering, cannot be anything other than remembering an object in its $U$, because the $U$ of an object is precisely one of the elements that describe the relationship between remembering consciousness and its object.

**Memory, consciousness and temporality**

The ideas we have expressed so far are part of a theory, the Memory, Consciousness and Temporality Theory (MCTT), that we have detailed elsewhere (Dalla Barba, 2002). Aspects of the MCTT relevant to the interpretation of confabulation are summarized below and schematically represented in Figure 11.3.

1. Events produce atemporal and aspecific modifications in the organism and, within the organism, in the brain. These modifications, represented in Figure 11.3 as X, Y, and Z, are atemporal in the sense that they do not contain any information concerning time. They do not represent the past, the present, or the future, nor are they organize according to the order of succession, i.e. there is nothing in Y, for example, that tells that Y come before Z and after X. They are aspecific in the sense that they do not contain any information specifying that they are representing episodes, meanings, rules, procedures, algorithms, etc.

2. The modifications in the brain can be more or less stable and more or less vulnerable depending on a number of variables. These variables include, among others, attention at encoding, emotional value of the event, depth of encoding, rehearsal and repeated experience of the same event, etc.

3. Consciousness means to be conscious of something in a specific way. That means that consciousness is not an aspecific dimension that passively receives and becomes aware of different types of already specified information, but rather that different types of consciousness exist, each representing an original and irreducible way of addressing the world. Different types of consciousness include, among others, TC and KC. TC means to become aware of something as part of a personal past, present, or future.
KC means to become aware of something as a meaning or as an element of impersonal knowledge or information.

(4) The object of consciousness represents a determination and an indetermination, what we have called U (uniqueness) and M (multiplicity). TC addresses the object’s U, whereas KC addresses its M.

(5) Less stable and more vulnerable patterns of modifications of the brain are necessary, but not sufficient, for the interaction between TC and the object’s U, whereas more stable and less vulnerable modifications of the brain are necessary, but not sufficient, for the interaction between KC and the object’s M.

**Temporal consciousness and confabulation**

How do the ideas that we have described apply to the interpretation of confabulation? Confabulation, as we have seen in the first part of this chapter, is
not limited to remembering personal past episodes, but also involves the patient’s present and future. According to the MCTT, in confabulating patients TC is still there, like in normal subjects, and addresses the object’s U, just like in normal subjects. These patients can still remember their past, they are present to a world, and can project themselves into a personal future. But in doing this they make errors, sometimes frankly bizarre errors. Actually, what is happening in these patients is that TC is still there but is not interacting with less stable patterns of modification of the brain. Most frequently the result of this condition is that personal habits and routines are considered in a personal temporal framework. When asked what they have done the previous day or what they are going to do the following day, confabulating patients typically answer with memories and plans that they usually have in their daily life. Although admitted to the hospital, they will say, for example, that the previous day they went out shopping and that the following day they will be visiting some friends, acts that presumably were part of their routine life. According to the MCTT, in this condition TC interacts with more stable patterns of modification of the brain and addresses the object’s M, habits, routines, repeated events, as U, a specific, unique past event.

Figure 11.4 Confabulation: TC addresses M as U.
It could be argued that patients who confabulate in episodic memory, orientation, and planning tasks are not necessarily conscious of a confabulatory past, present, and future but rather they simply produce the more plausible answer without having a subjective experience of remembering of being in that place at that time or of planning their future. If this were the case, our account of confabulation in the past, present, and future would be dismissed because TC wouldn’t play any role at all in confabulation. Confabulation would be just a sort of ‘best guess’ produced as a consequence of a faulty memory. Yet, there is evidence that patients who confabulate actually do become aware of their confabulatory past, present, and future. For example, patient MB (Dalla Barba, 1993) when asked to attribute a ‘remember’ or a ‘know’ judgement to his confabulations, he systematically gave ‘remember’ judgements. Also, the same patient showed he was ready to carry out his confabulatory plans (see also Baddeley and Wilson, 1986; Moscovitch, 1989; Burgess and McNeil, 1999; Metcalf et al., in press). In addition, from a clinical point of view, confabulating patients do not look like subjects who produce their ‘best guess’ in answering questions, but rather they seem to adhere completely to their confabulatory reports.

However, although habits, routines, and repeated events are the most frequent content of confabulation (Dalla Barba and Boissé, in preparation), confabulations are sometimes made of elements that cannot be traced back to any aspect of the patient’s history, present, or future situation. This is the case, for example, of patient MG who told the radiologist that he had accompanied a friend to the hospital and that the neurologist who was taking care of his friend realized that he also had neurological problems and so decided to refer him to the radiology department for a CT scan. Does the MCTT account for this case?

Well, it is clear that MG is immersed in a temporal existence: he remembers accompanying a friend to the hospital and he is there, about undergoing a CT scan because the neurologist noticed by chance that he had neurological problems. So TC is working in this case and, as it usually does, is addressing the object’s U, the specific past episode he remembers and the specific present situation he is involved in. Can we say in this case that TC interacts with more stable patterns of modification of the brain, like we said for habits, routines, and repeated events? Clearly there is no evidence in this sense. In this case, confabulation appears as being created ex nihilo. This is difficult to understand only if we are prisoners of the memory trace paradox and of the homunculus fallacy. MG’s confabulation is not a matter of failing to inhibit currently irrelevant memory traces, as Schnider believes (Schnider and Ptak, 1999). MG’s past and present world is not made of currently irrelevant memory traces. And it is not the result of the confusion between reality and imagination, like Johnson
would say, unless one wants to think that MG had once imagined the situation he is remembering and involved in, which is quite hard to believe. But if we consider that in MG TC is still there, the appearance of confabulation ex nihilo is not a problem, since we don’t need to trace confabulation back to an unconscious elsewhere, memory traces, imagination whatsoever. In fact, TC cannot escape to the general law of consciousness of being consciousness of something. So, if TC is there, it must be consciousness of something, i.e. it must temporalize its object according to the subordinate structures of temporality, past, present, and future. There is no need to ask where the object of this temporalization comes from because it doesn’t come from any sort of unconscious world. It is not behind consciousness, it is down there, right in front of consciousness. If we are not in a psychoanalytic perspective, do we ask where thoughts or imagined objects come from? We don’t. So why should we ask this question for TC? All we can say is that when TC temporalizes its objects, under normal conditions we can recognize a true temporal world, a real past, present, and future, probably because TC interacts with less stable patterns of modification of the brain. When TC temporalizes its object in confabulatory conditions, sometimes we can recognize habits, routines, and repeated events, probably because TC interacts with more stable patterns of modification of the brain, other times we don’t recognize any meaningful element of the patient’s history, present, and future situation.

So, the difference between normal memory and confabulation is that in the case of patients who confabulate we are faced with a confabulatory consciousness in which past, present, and future are somehow distorted. Distorted with respect to what? The truth, you will say; with what actually took place, is taking place, or will take place. But who establishes the truth? If temporality, and therefore its subordinate dimensions, is inherent to consciousness, what relationship of truth can there be between what you suppose to actually have taken place, to be taking place, or to be going to take place and the reality of TC? And yet for a witness, the doctor or the psychologist observing the patient who confabulates, his stories about the past, his perception of the present, and his future plans are confabulatory, in that they are totally at variance with what the listener expects as an answer, so much so that they push him to consider the patient’s answers as not normal, namely as confabulatory. Where does this discrepancy between what the observer expects and the answers he receives come from? In other words, what is the origin of this confabulatory temporality?

The answer to this question can only be a biological one. There is a biological difference between normal subjects and confabulators. This biological difference is represented by the brain lesion, which prevents the interaction between TC and less stable modifications of the brain.
It could be argued that we made it out with a shortcut since we have established an interaction between the biological level (patterns of modification of the brain) and the psychological level (TC). However, this is not a shortcut because TC depends on brain function and can be altered or abolished by a brain lesion. TC works normally in subjects without brain lesions, is altered in patients with brain lesions that cause confabulations, and is abolished, as we will see, in patients with brain lesions that cause amnesia. Now, the question is what are the brain structures which need to be intact for the normal functioning of TC, altered in confabulation, and damaged in amnesia? This is an experimental problem which needs more experimental work to be answered appropriately. However, the literatures on amnesia and on confabulation indicate that the hippocampus and the medial temporal lobe are good candidates as brain structures involved in the functioning and dysfunction of TC.

**Temporal consciousness and brain: Is the medial temporal lobe ‘temporal’?**

Patients with medial temporal lobe (MTL) lesions are known to be deeply amnesic for personal past episodes and to have relatively preserved general knowledge or semantic memory (see Kopelman, 2002 for a review). Patient HM was the first well-documented example of this condition (Milner, 1958). However, amnesic patients do not consciously remember their past, nor can they imagine their personal future. These patients are lost in a *non-time*, or in a sort of instantaneous present. The patient, N.N., described by Tulving (1985) is a good example of this condition. Tulving’s patient had preserved semantic abilities, i.e. N.N., but was unable to retrieve any episode from his personal past or to say anything about his future.

N.N. has no difficulties with the concept of chronological time. He knows the units of time and their relationships perfectly well, and he can accurately represent chronological time graphically. But in stark contrast to his abstract knowledge of time, his awareness of subjective time seems to be severely impaired. When asked what he did before coming to where he is now, or what he did on the previous day, he says that he does not know. When asked what he will be doing when he leaves ‘here’, or what he will be doing ‘tomorrow’, he says he does not know.

Here is part of the transcript of the interview with me as the interviewer:

E.T.: ‘Let’s try the question again about the future. What will you be doing tomorrow?’

(there is a 15-second pause.)

N.N. smiles faintly, then says: ‘I don’t know’.

E.T.: ‘Do you remember the question?’

N.N.: ‘About what I’ll be doing tomorrow?’
E.T: ‘Yes. How would you describe your state of mind when you try to think about it?’
(A 5-second pause.)
... When asked to compare his state of mind when he is trying to think about what he will be doing tomorrow with his state of mind when he thinks about what he did yesterday, he says it is the ‘same kind of blankness’.

(Tulving, 1985, p. 4)

As emerges from the description of this patient, what characterizes the core deficit of N.N. is a loss of TC. Past and future have disappeared as possible objects of his consciousness and the result is that N.N. is shut in an atemporal instantaneous present.

Further evidence of selective loss of TC in amnesia comes from patient DB described by Klein et al. (2002). DB suffered severe amnesia for the personally experienced past. By contrast, his knowledge of the non-personal past was relatively preserved. A similar pattern was evidenced in his ability to anticipate future events. Although DB had great difficulty what his experience might be like in the future, his capacity to anticipate issues and events in the public domain was comparable to that of neurologically healthy, age-matched controls. These findings show a dissociation between TC, which is impaired, and KC, which is preserved. Impairment of TC has also been documented in Alzheimer’s disease (Dalla Barba et al., 1999).

So, there is increasing converging evidence that lesions which produce amnesia also produce deficits of TC. This suggests that the integrity of the MTL and related structures is crucial for the normal functioning of TC. Further evidence on the involvement of the MTL in TC comes from a recent neuroimaging study which showed that in normal subjects the hippocampus was activated both when individuals remembered their past and when they imagined their future (Addis et al., 2007).

Now, if the integrity of the MTL is crucial for the functioning of TC, what happens in confabulation, where TC is present but functioning in an abnormal way? The most plausible answer to this question is that in confabulation the MTL is intact, but lesions to other brain structures prevent its normal functioning, which is the biological counterpart of the claim that in confabulation TC is preserved but no longer interacts with less stable modifications of the brain. This seems actually to be the case. Gilboa and Moscovitch (2002) found that only 2 out of 79 patients with confabulation had lesions involving the MTL. These patients had lesions in more than 20 brain regions, although the frontal lobe was preferentially affected, but almost all these patients had preserved MTL. Twenty-eight additional confabulating patients not included
in Gilboa’s and Moscovitch’s review also had preserved MTL (Dalla Barba, Cipolotti et al., 1990; Dalla Barba, 1993; Dalla Barba et al., 1998; Fotopoulou et al., 2004; Ciaramelli et al., 2006; Ciaramelli and Gheit, 2007; Fotopoulou et al., 2007). So, what these data show is that what confabulators have in common is not a specific lesion site but rather the integrity of the MTL, which is consistent with the idea that the MTL is essential for the function of normal and confabulatory TC. In this sense the MTL is ‘temporal’ because its integrity allows individuals to be consciously aware of a personal past, present, and future.

There are a number of important questions that we have not addressed because they were far beyond the aims of this work. Why confabulation is often a transitory symptom? Does confabulation reflect a disconnection of the MTL? Why amnesia, i.e. loss of TC, occurs for lesions outside the MTL? Is it also a matter of disconnection of the MTL? And if this is the case, what is the difference between confabulators and non-confabulating amnesics with preserved MTL? What is the exact role of frontal lesions in confabulation, if we reject, as we do, the possibility of unconscious monitoring? What is the exact nature of what we indicated as patterns of modification of the brain, if we reject, as we do, the notion of memory traces which would contain the past? Some of these are experimental questions that will be answered by experimental work. Others are theoretical and their discussion couldn’t have been included in these pages. What we have shown in these pages is that understanding confabulation is a difficult and ambitious enterprise. An enterprise voted to failure if based on theoretically untenable and experimentally undemonstrated explanatory idols like memory traces and unconscious monitoring. Confabulation is a conscious phenomenon, indeed a phenomenon that involves TC. A better understanding of TC, including its neurobiological correlates, is therefore demanded, legitimate and necessary.

References


REFERENCES


