Gianfranco Dalla Barba

Beyond the Memory-Trace Paradox and the Fallacy of the Homunculus
A Hypothesis Concerning the Relationship Between Memory, Consciousness and Temporality

Abstract: Most theories and models of memory are based on two assumptions that contain theoretical problems. These problems are reflected in the memory-trace paradox, which consists in believing that the past is contained in the memory trace, and in the fallacy of the homunculus, which consists in assuming the existence of an unconscious intentional subject. We will discuss these and present an alternative hypothesis concerning the relationship between memory, consciousness and temporality. This holds that consciousness is not a unitary dimension, but is the set of distinct and original modes to address the object. Among the modes of consciousness, a distinction is made between Knowing Consciousness (KC) and Temporal Consciousness (TC). KC describes the mode of addressing the object in order to know it. TC describes the mode of consciousness that temporalizes its object according the subordinate structures of temporality, the past, the present and the future. Finally it is shown how the hypothesis accounts for a variety of memory disorders and phenomena while avoiding the memory-trace paradox and the fallacy of the homunculus.

Introduction

In the last three decades research into the neuropsychology of memory has provided a considerable amount of data concerning memory dysfunctions in subjects with brain lesions. Both ‘negative’ symptoms, such as failure to recall or recognize some sort of information, and, more recently, ‘positive’ symptoms, like intrusion, false recognition and confabulation, have provided inferences on the organization and functioning of normal memory. A number of hypotheses and
ideas based on neuropsychological studies, and also on studies of normal subjects, have resulted in theories and functional models aimed at describing how memory works. Although each theory and model of memory is different from each every other in some respect, they can be roughly divided into two main categories: copy and reconstructive theories of memory. The first group of theories (e.g. Anderson & Bower, 1972; Kintsch, 1970; Morton et al., 1985; Ratcliff, 1978) consider memory of events as to be the reactivation of stored information as it was encoded. Reconstructive theories of memory (e.g. Jacoby et al., 1989; Johnson, 1991; Moscovitch, 1995; Moscovitch & Melo, 1997; Tulving, 1983) consider recovered memories as to be the result of the interaction between the stored information and a present cue.

As far as the retrieval of stored information and events is concerned, there is a more or less explicit agreement that it is the result of the activation of problem-solving-like processes that are engaged in operating an active search and selection in the long-term memory storage and in evaluating the result of their search and selection. In other words, memories are possible not only because information and events are stored, but also because some search and monitoring processes make their retrieval possible. This view is synthesized in models of memory that emphasize the role of search and monitoring processes in normal retrieval and their disruption in pathological retrieval (e.g. Baddeley & Wilson, 1986; Burgess & Shallice, 1996; Johnson, 1991; Moscovitch, 1989; 1995; Moscovitch & Melo, 1997).

Most of the current theories and models of memory, however, are based on two assumptions that contain theoretical problems which have never been clarified. The problems that these theories seem to have ignored can be summarized by the following questions: ‘What is retrieved?’ and ‘Who is controlling the retrieval?’ Concerning what is retrieved, the problem underlying most of the current theories is based on a paradox that we call the memory-trace paradox. Concerning who is controlling the retrieval, the theories are based on a fallacy, which we will refer to as the fallacy of the homunculus. In the first two sections of this article we will describe and discuss the memory-trace paradox and the fallacy of the homunculus. In the last two sections we will present a set of ideas concerning the relationship between memory, consciousness and temporality and show how they account for a variety of memory disorders and phenomena while avoiding the memory-trace paradox and the fallacy of the homunculus.

I: The Problem of the Past and the Memory-Trace Paradox

I. Past and memory traces

An assumption common to many old and new theories of memory is that of considering a memory as the result of the preservation of the past 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 1 depicts this point of view: event E is stored in the Long-Term Storage System as memory trace E'; the activation of E', with or without the involvement of a cue, results in E'', i.e., the conscious remembering of E.

![Figure 1](image)

*Figure 1. The relation between memory and consciousness in some current theories of memory*

This now classical way of thinking has 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 possibility of recollection on the preservation of an event inside a trace contains a paradox, fruit of a misleading assumption. The misleading assumption on which these theories are based is that time can exist in things. The past event which I now remember, for example my dinner yesterday with Paul, presents itself to my consciousness as a past event, in that as such, that is 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 Bergson says (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. As such, the objects of the
world are neither present, nor past, nor future, but they acquire a temporal dimen-
sion only in the presence of a person who goes to the trouble of making them tem-
poral. This erroneous assumption on which theories of memory are founded is
directly reflected in the paradox to which these theories fall victim.

2. 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 ner-
vous system) or virtual (the computable level of the functionalists) which I call
memory trace. What is the temporal nature of the modification which the event of
glass-on-the-table produces in the system, 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 is present, and that the event
represented by that modification is also 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 reinforcement of synap-
tic circuits, or of the particular codification of information, it will always be a
matter of present modification of something. What happens when that event con-
tained 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 reactiva-
tion of the modification that the event caused on a physiological, biochemical,
neuroanatomical, neurocybernetical 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 phenomena, perception, the conservation of the trace
and recollection, which do not reflect any notion of past.

![Figure 2. The memory trace paradox](image)
Figure 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 of memory overlook had already been clearly reported over fifty 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 never ceases 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 'me-ness' of memory of which Claparède (1911) speaks, the 'warmth and intimacy' with which James (1890) describes memory, or the 'pastness' of Bergson (1896), well describe certain characteristics of memory, 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.

3. 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 & Bower, 1974; Anderson & Bower, 1972; 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 is. The date or the ‘past’ information of the indicator, in themselves, tell us nothing, they are not past but they become so only if the past is already a 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. On the contrary, 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.

4. The order of succession

On the other hand, it would be useless 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. When I say that A comes before B, the anteriority of A presupposes in the nature of A an ‘incompleteness’ of A (instant or state) that points towards 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 din-
ner, 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 experi-
ence that that 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 epi-
sodes 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. As
stated earlier these theories also contain an illusion, the fallacy of the
homunculus, which deeply undermines their heuristic value.

II: The Homunculus Fallacy

1. 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 is therefore lost. But the
paradox of the memory trace is not the only problem with current theories of
memory. According to these theories (Baddeley & Wilson, 1986; Burgess &
Shallice, 1996; Conway & Tacchi, 1996; Johnson, 1991; Moscovitch, 1989;
Moscovitch & Melo, 1997), in order for the recollection of a memory or of knowl-
dge to be correct, that is for it to be the recollection of a memory or of informa-
tion 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.

Confabulation is a symptom which is sometimes found in amnesic patients and consists in involuntary production of ‘false memories’, that is the recollection of episodes which never actually happened, or which occurred in a different temporo-spatial context to that being referred to by the patient. For example, a patient who has been in hospital for several days may have the confabulatory memory of having done the shopping, or of having had dinner with the family the previous day. According to the hypothesis we have described, confabulation is the result of the dysfunctioning of memory-monitoring mechanisms, that is of the selection and verification mechanisms of the memory trace (Baddeley & Wilson, 1986; Burgess & Shallice, 1996; Conway & Tacchi, 1996; Johnson, 1991; Moscovitch, 1989; Moscovitch & Melo, 1997). The breakdown of these mechanisms is thought to impede 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, this hypothesis has never clarified whether the control-monitoring processes of the memory act on a conscious, voluntary basis or whether they are outside consciousness and inaccessible to it.¹

Let us look at this problem more closely. Supposing the monitoring processes assigned to the selection and verification of information were conscious and voluntary, in that case their action would be manifest to reflexive analysis since they would be operations of consciousness. The subjective experience of reflection would, then, show me that every time I am engaged in a recollection task, I

¹ Even though it is not explicitly stated, the same memory-monitoring mechanisms seem to be considered conscious by Burgess and Shallice and unconscious by Conway and Tacchi.
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 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 Wednesday at five in the afternoon, what I consciously begin is a search operation of successive approximation. I will remember, for instance, that on Friday morning I gave a talk in such and such a 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 trying to prepare 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 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, speaking in public, etc. Where I may hesitate is not on the content of my recollections, which can only present themselves apodeictically, 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 thematized past. 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.

One patient of ours (Dalla Barba, 1993b), for example, 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.
You could object that confabulation is 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. Here is an example. A patient of ours (Dalla Barba et al., 1997), MG, in hospital for a brain haemorrhage, while waiting for a CAT scan told the radiologist that he was in hospital because he had come to accompany a friend who was to be admitted to the neurology ward. Once they got to the ward, MG continued, the neurologist who was supposed to take care of his (non-existent) friend realized that there was something wrong with him too and so sent him for a CAT. The radiologist, who knew nothing of the situation, did not even suspect that at that moment MG was confabulating. In this case too, however plausible and indistinguishable from a true memory MG’s confabulation was, as demonstrated by the behaviour of the radiologist, one cannot consider it the result of the dysfunctioning of conscious selection and verification mechanisms. It is in fact 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, 1988; 1991; Johnson & Raye, 1981) 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.

2. Unconscious consciousness and intentionality
The subjective experience of reflection thus leads us to exclude the notion that 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 shall call the homunculus fallacy, that is the contradiction of postulating the existence of a type 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 man would contain another man, 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.

Let us first of all clarify what is meant by intentionality. I am conscious of this glass which is 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 con-
sciousness and intentionality, maintaining that there can be non-intentional con-
scious 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 behaviour. But why should a sudden feeling of
excitement be a state of non-intentional consciousness? Excitement, or its oppo-
site, 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 con-
sciousness is intentional in that it must, by nature, be consciousness of something.

3. Unconscious beliefs

Hypothesising the existence of intentional unconscious states involves the prob-
lem of supposing the existence of an unconscious intentional subject. The first
question to be asked clearly concerns the nature of the subject, this unconscious
intentional who. 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 uncon-
scious 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 them-
selves were complete and self-sufficient. In other words, that network of mean-
ings and beliefs which, according to Searle, ‘are in one’s head’.

Let’s 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 fulfil 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 behaviour 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 cognitivist 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, nonreflexive 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? It could be said that there is 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 excitation 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 won’t 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.

4. The problem of an unconscious ‘who’

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?

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’s 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: H\textsubscript{2}O. In H\textsubscript{2}O there is no liquidity, just as there is no solidity or gaseous state. In order for H\textsubscript{2}O to be liquid, something which has nothing whatsoever to do with its molecular structure is required. That is, H\textsubscript{2}O 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 H\textsubscript{2}O. 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 that present themselves to consciousness in accordance with different points of view. Water is both liquid and \( \text{H}_2\text{O} \) without one of these being cause of the other.

In short, in order to maintain that liquidity is an emergent property of \( \text{H}_2\text{O} \) 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 we 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.
III: A Hypothesis Concerning the Relationship Between
Memory, Consciousness and Temporality

1. Is a scientific theory of consciousness possible?

Up until now our discussion has dealt with present theories of memory which are based on a paradox, that is, on the idea that the past is passively preserved in memory traces. We have also seen that many of these theories are victim of a fallacy, the fallacy of the homunculus, that is, the assumption of the existence of a sort of unconscious consciousness in bad faith, the explanation of which sends one back to a solutio ad infinitum. The question we have to ask at this point is whether a scientific hypothesis on the relationship between memory and consciousness is possible without falling into the memory-trace paradox and into the fallacy of the homunculus.

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 its 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 objectivize 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. 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 objective ‘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.

2. Knowing consciousness and temporal consciousness

Earlier in this work 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 (for similar accounts, apart from the classic phenomenological tradition, see for example Chalmers, 1995; Varela, 1995). 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).

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 (Merleau-Ponty, 1945) or wesen ist 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, Tulving’s episodic memory (Tulving, 1983). But memories in consciousness of the past are not memories 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 past episode. In this case the episode’s being past is nothing but a notion, a quality that joins other qualities 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 emphasised 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 temporizes the world in the present form and that does not confuse itself with perception which is 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 possible in the future mode. By probable possible 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 possibles, 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 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 possible, but something that transcends the knowledge of the mundane being that I am.

3. Uniqueness and multiplicity.

One of the basic assumptions of our hypothesis is that consciousness must necessarily always be consciousness of ‘something’. 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. On the contrary, 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 a 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, are 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 a 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 a 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, what Tulving appropriately called episodic 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.

IV: Memory, Consciousness and Temporality

Before proceeding further to see the extent to which the hypothesis we have described can account for experimental and clinical observations, let us summarize its main points.

(1) Consciousness is not a unitary and generic dimension that passively receives what comes from outside, but is the set of distinct and original modes used to address an object. Among the modes of consciousness there is KC and TC. Each of these represents an original and irreducible mode of addressing the world.

(2) KC describes the mode of addressing the object in order to know it, that is in order to know its quality and quantity. KC roughly corresponds to what is generally referred to as semantic memory. Unlike semantic memory, however, KC does not presuppose any mental representation of the object or of its parts. As synthesis of past consciousnesses, KC is temporal, but it is not consciousness of time.

(3) TC is a differently organized and irreducible form of consciousness. TC is consciousness of time, i.e. it temporalizes its object according to the subordinate structures of temporality, the past, the present and the future. When it locates its object in the past, TC corresponds to what is referred to as episodic memory. However, consciousness of the past distinguishes itself from episodic memory in that it does not presuppose the existence of episodic-memory traces. TC establishes itself in KC and presupposes KC.

(4) Since consciousness must be consciousness of, every type of consciousness is consciousness of an object. The object of consciousness represents a determination and an indetermination, a U and a M. However, the U of the object is founded in its M. The relationship between M and U varies according to the mode of consciousness in which the object is addressed. The balance between M and U moves towards U when the object is addressed in the mode of TC and towards M when it is addressed in the mode of KC.

As far as the neural correlates of TC and KC are concerned, two further assumptions can be made:
(5) Since TC and KC are distinct and original modes of consciousness, it is possible that different patterns of neural organization and activity correspond to TC and KC. However, since TC is established in KC, the different patterns of neural organization and activity should largely overlap.

(6) Since TC and KC are the constant synthesis of what they have been, it is possible that the patterns of neural organization and activation corresponding to TC and KC are in constant change. The rate of change of neural patterns of organization and activation should be a function of the rate of unicity of experience, i.e. the ratio between TC and KC. Accordingly, patterns corresponding to TC, i.e. U, should be less stable than patterns corresponding to KC, because TC is founded in KC. This would be reflected by a greater instability and vulnerability of TC compared to KC, that is of U compared to M. Assumptions 1–6 are summarized schematically in the model given in Figure 3.

![Figure 3](image)

*Figure 3. A hypothesis concerning the relation between memory, consciousness and temporality.*

NS = nervous system; X and Y = patterns of organization and activation of the nervous system; TC = temporal consciousness; KC = knowing consciousness; U = unicity; M = multiplicity.

As far as pathology is concerned the following predictions can be made:

(1) Since TC and KC are distinct, organized and irreducible forms of consciousness, they should also be distinct also in pathology, i.e. it should be possible to observe specific deficits that reflect the mutual relationship of TC and KC.

(2) However, since TC is founded in KC, it should not be possible to observe deficits of KC without deficits of TC. On the contrary, it should be possible to observe deficits of TC without deficits of KC.
(3) A deficit of TC should impair, although to different degrees, all three dimensions of temporality. In other words a patient with an episodic-memory deficit is also likely to have a deficit of consciousness of the personal present and future.

(4) In phylogeny and ontogeny the development of KC should precede that of TC. On the contrary in progressive degradation of consciousness, the deficit of TC should precede that of KC.

(5) Since M and U of the object of consciousness are functions of the modes of consciousness, it should be possible to observe cases in which TC sees M as U and cases in which KC sees U as M.

We will now demonstrate to what extent some of the predictions of our hypothesis are confirmed by clinical and experimental observations.

The first prediction of our hypothesis concerned the possibility of observing selective deficits of TC without any comparable deficits of KC. We claim that amnesia reflects this very condition. Amnesic patients are described as reflecting a deficit of episodic memory with relatively spared semantic abilities (e.g. Cermak & O’Connor, 1983; Dalla Barba, 1993a; Dalla Barba, Cipolotti & Denes, 1990; Kinsbourne & Wood, 1975). However, an episodic-memory deficit does not exhaustively describe what happens in amnesia. In fact, amnesic patients not only do not remember their personal past, that is they can no longer locate the object in the past temporal mode of consciousness, but neither can they locate their object in the present or in the future, as stated in prediction 3. 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. KC, 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.)

N.N.: ‘Blank. I guess.’
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 the inability to access his personal past and future. 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.

Evidence for selective deficits of TC does not come only from amnesia but also from another type of memory impairment, i.e. confabulation. We have recently described the case of a patient, G.A (Dalla Barba, Cappelletti, Signorini & Denes, 1997), in whom confabulation involved episodic memory as well as her performance in tasks of time and place orientation and in tasks where she was required to plan her personal future. For example, to the question ‘What are you going to do in a few minutes?’, G.A. answered ‘I am going home to cook the supper’. Since her disease, she had actually never cooked and she was living one and a half hours from the hospital. To the question ‘What are you going to do tomorrow?’ she replied ‘I am going out shopping alone by car’. She had never done so since her disease and she would certainly not do so on the following day. In another study (Dalla Barba et al., 1999), it was shown that patients with Alzheimer’s disease confabulated when they were required to retrieve a personal episode and, although less frequently, when they were required to make a personal plan. So there is at least preliminary evidence that a deficit of TC affects each of the subordinate structures of temporality, although probably to a different extent. Further research will help to confirm whether a deficit of TC is a systematic finding in patients described as suffering from an episodic-memory deficit.

Prediction 2 states that, since TC is founded in KC, it should not be possible to observe deficits of KC without a deficit of TC, because this would reflect a selective deficit of KC. Several authors have argued for the existence of selective impairments of semantic memory without any comparable deficits of episodic memory. However, in reported cases of selective semantic impairment, episodic memory is claimed to be normal on the basis of the observation that day-to-day memory is preserved (e.g. De Renzi et al., 1987; Hodges et al., 1992; Hodges et al., 1994). Typically these patients are impaired in a variety of semantic-memory tasks, but they usually remember day-to-day events including the different testing sessions and their content. These observations might lead to the conclusion that episodic memory is independent of semantic memory, and that TC is independent of KC, and that a deficit of KC can occur without any comparable deficit of TC. However, the comparison of patients’ performance in day-to-day memory and in formal semantic-memory testing can hardly lead to firm conclusions on the relationship between semantic and episodic memory. In fact, day-to-day memory is observed in ecological conditions, which may be an easier situation than that of formal semantic-memory testing, and variables that affect one situation might be different from variables that affect the other. In addition, patients with selective semantic-memory deficits are usually severely impaired in learning tasks which are traditionally considered to tap episodic memory, (e.g. Basso, Capitani &
Laiacona, 1988; De Renzi et al., 1987; Hodges et al., 1994; Pietrini et al., 1988; Warrington & Shallice, 1984; Wilkins & Baddeley, 1978).

Prediction 4 of our hypothesis states that in phylogeny and ontogeny the development of KC should precede that of TC. In contrast, in progressive degradation of consciousness, the deficit of TC should precede that of KC. As far as the first point is concerned there are no data that can confirm or disconfirm this prediction. As far as ontogeny is concerned, however, evidence in favour of the hypothesis relies on the observation that in infancy and childhood the development of semantic abilities, KC, precedes the development of TC. As far as the second point is concerned, favourable, partial evidence is to be found in the study of Alzheimer’s patients. In these patients, in fact, an episodic-memory deficit systematically precedes a semantic memory deficit.

Prediction 5 states that since M and U of the object of consciousness are functions of the modes of consciousness, it should be possible to observe cases in which TC sees M as U and cases in which KC sees U as M. As far as the first point is concerned, indirect evidence comes from the study of confabulation. In fact, confabulation in episodic memory (e.g. Dalla Barba, 1993a; Dalla Barba et al., 1999; Dalla Barba et al., 1997; Dalla Barba et al., 1990), in orientation in time and place and in personal future planning tasks, as in GA’s case, can be interpreted as the result of a condition in which TC is still there but, in order to remember the personal past, to be oriented in the present and to plan a personal future, addresses M as U. This condition is schematized in Figure 4. In other terms, this is a condition in which TC is still working but is no longer able to accomplish its usual task, i.e. to operate a sort of fine grain search in order to set up a personal temporal workspace. What TC does instead in this condition is to address M of the object as U, so that as a result habits or personal semantic information are considered in a personal temporal framework. When asked what they did the previous day or

Figure 4. Confabulation: TC addresses M as U; Implicit memory: KC addresses U as M.
what they are going to do the following day, confabulating patients of this type typically answer reporting as memories and plans what they usually do in their daily life. Although hospitalized, they will say, for example, that on the previous day they went out shopping and that the following day they will be visiting friends, acts that presumably were part of their routine life.

It could be argued that patients who confabulate in episodic memory, orientation and personal future planning tasks are not necessarily conscious of a confabulatory past, present and future but just produce the more plausible answer without having the subjective experience of remembering, of being in that place at that time or of planning their actions. If this were the case our account of confabulation in the past, present and future would be dismissed because TC would not play any role at all in confabulation. Yet there is evidence that patients who confabulate actually do become aware of their confabulatory past, present and future. Patient M.B., for example (Dalla Barba, 1993a), when asked, following the procedure described by Tulving (1985) to attribute a ‘remember’ or a ‘know’ judgement to his confabulation in episodic-memory tasks, systematically gave ‘remember’ judgements. Also, the same patient showed himself ready to carry out his confabulatory plans. On one occasion, for example, he said that he was looking forward to the end of the testing session because he had to go to the shops to buy some new clothes. On that occasion he actually attempted to leave his hospital room, claiming that there was a taxi waiting for him downstairs (see also Baddeley & Wilson, 1986; Moscovitch, 1989). In addition, from a clinical point of view, confabulating patients do not look like subjects who, in answer to questions, produce their best guess, but they seem to adhere completely to their confabulatory reports. Therefore, although more experimental evidence is necessary, it doesn’t seem to be misleading to consider confabulation in episodic memory, orientation and future planning tasks as a condition in which TC addresses M as U.

As far as the second point of prediction 5 is concerned, i.e. the possibility that KC addresses U as M, indirect evidence in favour of this comes from data concerning implicit memory. In fact, what is indicated as implicit expression of memory in amnesic patients, can be viewed as the reflection of KC that sees as M the U of the object. In other words, past experience is not seen by amnesics in its U, as it would be if it were seen in the mode of TC, but as M, i.e. in the mode of KC. Claparède’s amnesic patient, who, having been pricked by a needle hidden in Claparède’s hand, refuses to shake his hand a second time because ‘sometimes there are needles hidden between the fingers of hands people shake’ translates in an impersonal generalization, i.e. in a M, the U of the object, i.e. the past experience of being pricked. The amnesic patient who better executes a given task the fifth time compared to the first, while believing he is performing the task for the first time, is proof of acquired ‘knowledge’ of which there is no TC, if not that which reflects the relationship between KC and the object seen in its M instead of in its U.

The hypothesis we have described accounts for a variety of memory disorders and phenomena. Further theoretical and experimental work will show whether
this hypothesis is to be confirmed and if it can be developed and organized into a more articulated theory on the relation between memory, knowledge and consciousness. At present, however, the hypothesis, though probably incomplete, at least has the advantage of accounting for certain memory phenomena without falling into the memory-trace paradox or the homunculus fallacy.

References

Claparède, E. (1911), ‘Recognition et moïté’, Archives de Psychologie, 11, pp. 79–90.


Sartre, J-P. (1943), L’être et le néant (Paris: Gallimard).


Paper received June 1999