**A critique of**

**the causal theory of memory**

**Marina Trakas**

**Director: Jérôme Dokic**

**Philmaster (EHESS)**

**2010**

**ACKNOWLEDGEMENTS**

I would like to thank first of all, my mother, Olga Mattiuzzo and my father, Antanas Trakas, for everything they did and do for me, but especially for teaching me how to love.

Second, Rémi Pelletier, for selflessly devoting to me a great deal of time; also to his wife, Marie-Claude Gordi, and his sister, Anne Pelletier, for loving me like a member of their family. With you I feel safe in the North.

Third, Jérôme Dokic, for trusting me during these two years even if I am not sure that I deserve it.

Fourthly, the Universidad Nacional de La Plata which educated me since I was four years old absolutely for free, especially Oscar Esquisabel, one of the best professors I had there.

Fifthly, to Kirk Michaelian, for helping me with the bibliography and the linguistic correction.

Finally, I also would like to thank to Nhora Patricia Palacios Trujillo, Belén Neira, Julia Zaparart, Cintia Indarrimendi and Germán Oyarzún Silva for all the support, love and care we mutually gave each other during the beautiful time that we spent together in Paris.

**CONTENTS**

**Introduction 1**

**Chapter 1: The causal theory of memory 3**

The traditional causal theory of memory 3

A new reformulation of the causal theory of memory: Bernecker’s account 13

Typology 14

A preliminary analysis of the conditions of memory 20

The memory trace 21

Arguments in favor of a causal theory 25

The object of memory 32

Truth conditions of memory 36

Similarity of content 36

Some concluding remarks 42

Going a little (but not too much) beyond the traditional causal theory of memory 44

Matthen’s attempt: New content but same old trace 45

Michaelian’s account: A trace that changes but just a little 47

General problems of the causal theory of memory 51

**Chapter 2: Psychological models of memory 57**

The spatial analogy 57

Symbolic models 58

Quillian’s model 60

The ACT theory 64

Connectionists models 78

McClelland’s model: local connectionism 79

McClelland and Rumelhart’s distributed connectionist model: 82

**Conclusions 96**

**Bibliography 99**

**INTRODUCTION**

In the last fifty or sixty years there were two currents in philosophy of memory that tried not only to analyze the concept of memory but also to explain what is a memory event[[1]](#footnote-1) and what kind of relation it has to the past and how. The difference between these two lines of thought is centered on the acceptance or rejection of the idea of a memory trace in the mind or brain[[2]](#footnote-2) of the subject, which on one hand is produced by certain past experience and on the other hand produces the memory of it. Those who are in favor of this notion maintain a causal theory of memory: the past experience[[3]](#footnote-3) is causally connected to the present memory through the persistence of a memory trace. Those who are against this notion, either analyze the memory event in terms of retention of knowledge or a capacity to represent the past experience, or give a causal explanation of the memory directly from the past experience. I will reserve the expression “causal theory of memory” exclusively for the first current.

Nonetheless, the second current does not seem to be grounded; first, as I will develop later, if we want to explain how the retention or the capacity to represent the past experience is possible, we must invoke something that remains in the cognitive system and whose nature couldn’t be very different from that of the memory trace; second, not only it is assumed that past events exert a causal action through time and thus have some sort of existence but, what is more important, causality is analyzed exclusively on a metaphysical or epistemological level, an analysis that rules out any possibility of considering what happens on a cognitive level. This makes it impossible to take into consideration any kind of psychological, neurophysiological or even philosophical analysis of the mind.

Because I believe in the possibility of integrating different disciplinary approaches in order to explain the memory phenomenon in a more accurate way, I decided to rule out this philosophical account (even if I will occasionally refer to it) and focus my attention exclusively on the causal theory of memory, the only philosophical theory that, thus far, through the notion of memory traces, seems more plausible and can be compatible and integrated with some psychological approaches to memory.

In general terms, the causal theory of memory assumes two fundamental postulates:

(a) a causal relationship between the past experience and the current memory representation via a memory trace, which implies a univocal correspondence between a past experience, the memory trace and the memory event of it;

(b) a similarity of content or structure between both representations (including in some cases the memory trace);

Nonetheless, as I will try to show, the causal theory presents some important conceptual problems and in some cases makes assumptions that luck justification. Consequently, in the first part on this essay I will present a number of different versions of the causal theory of memory that have been proposed in the literature and will analyze to what extent they actually explain memory phenomena. In the second part, my attention will be focused on the psychological models that have been developed to explain the same topic, in order to see if they suggest the same conception of memory proposed by the causal theories or instead provide a glimpse of another possible analysis of memory phenomena. Therefore, I will concentrate on the analysis of theories and will unfortunately omit empirical studies about memory (due especially to the limitations of space). Finally, I will conclude that the causal theory in all the different versions developed so far does not constitute a good philosophical account of memory. In the last part, I will outline some points that an alternative philosophical theory of memory should take into account.

**CHAPTER 1: THE CAUSAL THEORY OF MEMORY**

**The traditional causal theory of memory**

Though the notion of memory traces has a long tradition, I will begin with the article “Remembering” written by Martin and Deutscher in 1966, because of the influence it has had and continues to have as the paradigm of a causal theory of memory based on traces, not only for their supporters but also for their detractors.

In this article, Martin and Deutscher define the necessary and sufficient conditions for a memory event in order to differentiate it conceptually from others mental states. The criterion they propose is the following one:

“ 1. Within certain limits of accuracy he [the subject] represents that past thing.

2. If the thing was 'public', then he observed what he now represents. If the thing was 'private', then it was his.

3. His past experience of the thing was operative in producing a state or successive states in him finally operative in producing his representation.”[[4]](#footnote-4)

Concerning clause 2, it is not clear what Martin and Deutscher want to distinguish. If we take into consideration the wording of the clause, the distinction could refer to the difference between experiences and mental states, an interpretation that could be reinforced from a later reference to “public events” and “private episodes”[[5]](#footnote-5), where “public events” could include not only personal experiences but also facts learned, because the facts also have to be represented by the subject in order for it later to be possible for him to remember them[[6]](#footnote-6). But the lack of other references plus the fact that all the examples refer to episodes experienced by the subject do not really suffice to show that the interpretation of the distinction mentioned is the right one. We could also think that it simply refers to the more traditional classification between episodes and facts, which would correspond to the distinction that Martin and Deutscher draw at the beginning of the article between remembering that1, where the representation of the event is not due to the experience of the event happening, and remembering that2 where the subject has to have experienced or observed the event directly.

In relation to clause 3, we would like to make two things clear. First, the term “operative” is defined by Martin and Deutscher as the traditional conception of “cause”, that is, as a causal condition which may be necessary or sufficient but need not be either. It could be sufficient, but generally it is not because “innumerable other factors and events are always necessary for representation of a past observation to occur”[[7]](#footnote-7) (external stimuli, specific conditions, etc). It could be necessary for producing the representation of the past event in some cases, but not in other cases, because it “could have been obtained at the same time by another method, had the operative condition not been present”[[8]](#footnote-8) (as in the first example cited on the next page).

Second, Martin and Deutscher subdivide this third condition in three subclauses:

a. the first emphasizes the “operativeness” of the experience in producing the memory event;

b. the second warns us about cases where prompting is operative for the memory representation: for a mental state to be an act of memory, the prompting cannot be complete or happened in his past than was supplied by the prompting[[9]](#footnote-9);

c. the third clause adds the isomorphic condition: “ The state or set of states produced by the past experience must constitute a structural analogue of the thing remembered, to the extent to which he can accurately represent the thing”[[10]](#footnote-10).

These three conditions, i.e., past perception or introspection, accuracy and a causal connection, only concern declarative memory. Concerning the third condition, it assumes the existence of a trace which, according to Martin and Deutscher, constitutes the only guarantee of the operativeness of the past experience in producing the memory event, on the one hand, and the structural analogy between them, on the other hand.

Even if the content of their paper is much richer than the general ideas just described, since my main interest here is in memory traces, I’m going to focus only on these two last points, points which characterize a tendency in philosophical theories of memory that continues to prevail today, as we will see later.

Let’s start with the first point: the causal relation between the past experience and the memory of this past experience. There are two kind of examples mentioned in the article that are very illustrative of this relation, each one of different aspects.

For the first case, there are distinct variants that nevertheless express more or less the same idea; thus I will develop the general schema that underlies all of them. A person A experiences something x and describes this experience (with all the details included) to a friend B. After that, A forgets x (due to an accident for example) but B tells him all the details concerning x in such a way that A seems to remember his own experience again. In the course of time, A remembers x but forgets about the source of his memory: the story told by his friend. He wrongly believes that the source of his memory of x is his own experience of x. Because A cannot add any details to his experience beyond what B told him, Martin and Deutscher conclude that even if A believes that he is remembering x, he is not actually remembering x, because his past experience of x is not operative in producing this memory, as clause 3a requires (in fact, it is the past experience of his friend describing to him the experience x which constitutes the real cause of his memory).

If the first examples refer to the cases where someone is not actually remembering or retrieving the past experience which he believes he is remembering, the second example refers, in contrast, to the opposite case: someone who actually remembers without knowing or believing that he is remembering. Martin and Deutscher invite us to imagine a painter who thinks that he paints an imaginary scene, e.g., a farmyard, when he is actually representing a scene that he saw in his childhood. His parents can give a testimony of this, because they remember their son seeing the farmyard. Given the evidence provided by other people, Martin and Deutscher conclude that the painter is not imagining, as he thinks, but he is actually remembering without knowing that he is remembering[[11]](#footnote-11).

These two thought experiments show that what really defines not only a memory event but also an act of remembering is the existence of an objective causal relation between the past experience and the current representation of it, independently of the subjective belief that one is remembering (what Tulving, in the vocabulary of cognitive psychology calls “autonoetic consciousness”[[12]](#footnote-12)) or a feeling of familiarity or pastness that for example is indispensable in some accounts of memory like that one of Russell[[13]](#footnote-13). However, because Martin and Deutscher defend the classical notion of causality as locally constrained, they cannot admit the possibility that the past experience can cause a memory event directly over a spatio-temporal gap; that’s why they need to introduce the idea of a persistent medium that makes the causality effective: the memory trace. Consequently, the memory trace is the only medium through which the past experience can be operative in producing a current representation. Also, as Martin and Deutscher remark, the memory trace is the only way to ensure that only what we have experienced or learned in the past can cause a current representation in us. In Martin and Deutscher’s account, the idea of a memory trace is a necessary implication of a causal theory of memory. However, even if this concept is presented as a key concept, it is not well developed by the authors. They do not specify the following points:

a. if the trace is a theoretical construct and/or a physiological change in the brain;

b. what kind of causal role the trace plays in producing the act of remembering given that the past experience is considered to be the principal operative factor.

We can nevertheless infer possible answers to these questions from certain paragraphs of the article: although the trace is a theoretical implication of the causal theory of memory, it is also a physiological reality, and its causal role would be more like a medium that enables the conservation of the causal bond between the past experience and the act of remembering it. This last point could questioned, because in clause 3(a) the authors explicitly say that “past experience of the thing was operative in producing a state or successive states in him finally operative in producing his representation”[[14]](#footnote-14). But even if it appears at first that the trace is simply a consequence of the past experience which, once this experience has disappeared, turns into the only operative factor which causes the act of remembering, after reading Martin and Deutscher’s article we can note the relevance that they give to the past experience in producing the act of remembering and clearly not to the trace (the concept of which, as I have noted, it is not even developed in detail in the article). Thus I maintain that in Martin and Deutscher’s account the trace is a medium which conserves the causality over the time rather than an entity which, even if it has a cause, it is independent enough of it to be able to cause something else itself (which would only allude to the transitivity property of causality)[[15]](#footnote-15). The general schema corresponding to this idea is the following one:

(schema 1) experience of x memory of x

trace of x

which can be read as: “A past experience of x is operative in producing the memory of x through a memory trace of x.”

And not this one:

(schema 2) experience of x trace of x memory of x

which would mean “A past experience of x is the cause of a trace of x, and the trace of x is the cause of the memory of x” (transitivity of causality).

This interpretation is reinforced by the distinction drawn by Martin and Deutscher between being “operative for” and being “operative for and in”, even if this distinction is outlined in relation to clause 3b, in order to distinguish where there is a prompting, real cases of remembering from false ones. “Operative for” is defined as:

“E (experience) being operative for the circumstance P (prompting) which helps to bring about R (representation).”

whereas “operative for and in” is described as:

“E (experience) being operative in the circumstance P (prompting) which helps to bring about R (representation).”[[16]](#footnote-16)

In the first case, the experience E produces the prompting P and P produces the memory representation R. Even if in one sense we could say that E produces R by transitivity, P’s production of R is independent of what brought about P (it could be an experience different from E). This would be precisely the case of the first example schematized above, that is, of a person who, at a first time has an experience E, tells it to a friend and forgets it due to an accident. At a later time, his friend tells him about the experience (the prompting P), then he has an accident again and forgets the fact that his friend told him the story, but after that he represents the experience (R) and thinks he is actually remembering the experience E. This case would not be a case of remembering because the experience E is not operative for producing R.

In the second case as well, the experience E produces the prompting P and P produces the memory representation R. But the experience E also produces another factor which acts with P to bring about R: the memory trace T. So in this case, given T, the experience E via T is still relevant for the production of the representation R. Only cases where the prompting in conjunction or simultaneously with the trace produces the mental representation are real cases of remembering.[[17]](#footnote-17)

Why does this difference constitute reason in favor of our interpretation of the role of the trace as a vehicle of the causality in Martin and Deutscher’s account? Because if the past experience when there is prompting, is only considered relevant for the production of a current representation via the production of a trace and not through the production of any other fact that could also bring about the same representation (and in an independent way), it seems likely that in general terms the trace is considered not as an entity independent of the past experience but as a kind of prolongation of it and of its causal role, as described in schema 1.

Furthermore, this interpretation is reinforced by the second point that I would like to develop, namely, the structural analogy between the past experience and the memory event. I will first develop this idea; below, I explain the sense in which it contributes to my interpretation of the role that the trace plays in Martin and Deutscher’s account.

As I noted above, there is not only a causal relation between the past experience and the memory event but also structural analogy between them. I suggest that the concept of analogy, which, like the concept of “trace”, is crucial to Martin and Deutscher’s causal account, is not better developed than the latter concept. It seems that the analogy is double: between the past experience, the trace and the memory representation. From the article, we can only infer the following features of the analogy:

a. it is structural, although the idea of a structural analogy is far from clear: which is the structure that is analogous between the past experience, the trace (which would have a physiological basis) and the memory event? Does it refer to a semantic structure, that is to the content of the past event, the trace and the memory event? But the idea that the trace has a semantic content would entail that is a representation; does this make the memory event a representation of second degree (like works of art in Plato)? It would be even more difficult to determinate this we introduced the idea outlined in the clause 3(a) of an analogy with the thing remembered, which would be different from the past experience and in certain way the cause of it (at least in the cases of public events).

b. it is not perfect, which means that it does not have to mirror all the features (the authors say that the notion of “all the features of a thing” makes no sense). It only has to contain “at least as many features as there are details which a given person can relate about something he has experienced”[[18]](#footnote-18). But once more, not only does the idea of a perfect analogue not make sense but the idea of having a kind of quantitative equivalence between the “features” of the experience or the thing and the “features” of the memory representation also does not mean anything, as Malcolm pointed out[[19]](#footnote-19).

Even if the notion of structural analogy is very poorly developed in Martin and Deutscher’s account, the idea of incorporating it counts as a second point in favor of schema 1 in our interpretation of the role of the trace. Why? Because the trace does not have a structure (or content) different from that of its cause (the past experience), but simply reproduces it. More specifically, here again we can consider the trace as a simple vehicle, a simple medium whose function is to conserve the structure of the past experience in order to transmit this same structure to the memory event. This idea of a trace as a simple vehicle of structure is clearly more compatible with the idea of a trace as a vehicle of causality which is represented in schema 1 than with the idea of a trace as an autonomous entity that by itself produces an act of remembering (schema 2). In fact, the autonomy of the trace suggests the idea of a trace whose structure or content could change, possibility that is not admitted by Martin and Deutscher.

This schema can be extended if we take into consideration another point suggested in the article. Recall the first criterion, which requires accuracy in the initial representation of the experience. Taking into consideration the distinction between public events and private episodes, when the experience is a private episode, that is, an introspection of the subject’s own mental states, talking about accuracy does not make very much sense. But when the thing or event we represent is public, the requirement of accuracy implies that the person correctly represents the thing or event. In Martin and Deutscher’s words: “We have agreed that when the person says that he saw a man, his memory is not at fault, though he actually saw a wax dummy and not a man. But though his memory is not at fault, he cannot be said to remember seeing a man at that time to be remembered. Our criterion requires only that he represent correctly what he does remember”.[[20]](#footnote-20) This means that for the cases of public events, the schema could be reformulated as it follows:

(schema 1a) x representation of x memory of x

trace of x

This can be read as: “A public event x causes an internal representation of x which is a structural analogue of x; this internal representation of x through a trace which is also a structural analogue causes a memory of x, which conserves the same structure”. We have to remember that the causality here is understood in terms of “operativeness”, at least as far as the bond between the past representation and the memory of it concerned[[21]](#footnote-21). This important point seems to be forgotten by Bernecker when he interprets Martin and Deutscher’s notion of causality in counterfactual terms:

“1. If S had not represented that p in the past, he would not represent that p now.

2. S’s past representing that p in the past is a causally necessary for his present presenting that p.”[[22]](#footnote-22)

Even if this idea can be inferred from a short passage mentioned in the article[[23]](#footnote-23), Martin and Deutscher remark more than once that the notion of an operative cause is not equivalent to the notion of a necessary cause, so that the actual representation of *p* can be caused by means other than the past representation of p. In this case, clearly, the person would not be remembering, because, remembering to occur, the past representation of *p* must be one of the efficient causes; but this does not mean that the person could not represent *p* with accuracy (remember the first though experiment that we mentioned: a person who represents an experience of his own past not because he represented it in the past, but because a friend described his past experience to him, though he does not remember that). The past experience of *x* is only causally necessary for producing an actual representation of *x* in real cases of remembering. So, we can reformulate Bernecker’s interpretation in the following way:

1. If S had not represented that p in the past, he would not remember that p now (even if he could represent that p).

2. S’s past representing that p in the past is causally necessary for his present memory of p.

On Bernecker’s interpretation, Martin and Deutscher’s argument fails establish the necessity of causation for memory, because the dependency expressed by a counterfactual does not necessarily have to be causal dependency: “Even if one’s present retelling of a past event is counterfactually dependent on one’s previous having witnessed that event, this doesn’t have to mean that the witnessing is the cause of the retelling. The sort of dependency expressed by counterfactuals is considerably broader than strictly causal dependency and causal dependency is only one among the heterogeneous group of dependency relationships that can be expressed by counterfactuals”.[[24]](#footnote-24) On my interpretation, Martin and Deutscher are not trying to give an argument to demonstrate the necessity of causation for memory, but are simply trying to define the concept of “memory”, to discover its intrinsic nature; in their account, the necessary bond between a memory of x and the past representation of x is true by definition. The problem is that this definition is not operative (in the common sense of the word and not in Martin and Deutscher’s special use) for recognizing a real memory event; however, as Martin and Deutscher point out, this is not a negative point against their account but a positive one: “we will not be able to decide whether the person is remembering, but this is no objection to our theory, but the very reverse. One of the faults of analyses of memory which have recently been given is that they leave insufficient room for the real doubtswhich canarise concerning whether or not someone is remembering something. It is a common fact that there can be doubt whether a person who gives a true account of his own past is remembering. For instance, as pointed out earlier, we may fail to have enough evidence to decide whether someone who gives a true”.[[25]](#footnote-25)

In order to summarize the principal ideas formulated in Martin and Deutscher’s account, we can restate the following claims:

(1) An accurate and correct past representation for the cases of public events (which include episodes and facts);

(2) A causal relation (defined in terms of operativeness) between the past representation[[26]](#footnote-26) and the memory event generated by a memory trace (in the case of a public events, we have to add in the first place the causal relation between the external thing or event and the representation of it);

(3) An univocal correspondence between a past representation and a memory event;

(4) An structural analogy between the past representation, the memory trace of it, and the memory event.

The last three clauses (2), (3) and (4) imply that:

(a) there is a biunivocal correspondence between a past representation and the memory trace of it and a univocal correspondence between this memory trace and the memory event;

(b) the memory trace is a single and independent unit that survives unchanged over time;

(c) even if over time there can be many memory events stemming from a past representation, all of these will be similar as far as their content is concerned, because they will have an analogous structures. There is only reproduction of content and neither transformation or a generation.

**A new reformulation of the causal theory of memory: Bernecker’s account**

As I noted at the outset, the causal theory of traces, even if it was subjected to numerous critiques, prevailed against other accounts of memory, perhaps due to its being the most intuitive of the available accounts. Not only is the theory implicitly accepted in other areas of philosophy of mind, but it also been defended in an updated form by Sven Bernecker in his first book *The* *Metaphysics of Memory* (2008), as well as in his more recent book *Memory: A Philosophical Study* (2010). I will analyze the theory of memory developed in the first book[[27]](#footnote-27), not only because it is a recent update of the causal theory of traces but also because it is the only recent philosophical study concerned with memory.

Four key ideas are developed in the book: a causal theory of memory, a direct realism about memory, an externalist response to skepticism about memory knowledge and a contextualist account of the factivity constraint on memory. Once more, I will focus my attention on the two aspects of the theory that are of interest to me here, that is, the causal relation and the content relation between the past experience, the trace and the memory event[[28]](#footnote-28), in order to determine to what extent Bernecker’s account differs from Martin and Deutscher’s proposal. But I want to begin with some (fairly lengthy) remarks concerning the kind of memory that is the object of Bernecker’s analysis.

*Typology*

Bernecker refuses the traditional distinction between episodic memory, semantic memory and procedural memory, because he claims (without, however, offering evidence in support of the claim) that it is impossible to establish a clear distinction between the two types of declarative memory. The new classification that he introduces is a grammatical one, based on the different kinds of objects that go with the verb “remember”. According to this criterion, “one can remember (i) objects (e.g., remembering Fido), (ii) properties (e.g., remembering Fido’s floppy ears), (iii) events (e.g., remembering Fido biting the mailman) and (iv) facts (e.g., remember that Fido bit the mailman)”.[[29]](#footnote-29) Object-, property- and event memory are considered to be non-conceptual, whereas fact memory would be the only type that requires the possession of concepts (which is absolutely questionable). As Bernecker recognizes, fact memory is similar to what Russell described as memory by description, and object-, property- and event memory are similar to memory by acquaintance. In fact, other philosophers had already divided memory according to the nature of the object remembered, including, for example, Broad, who distinguishes four kinds of memory: memories of things, of events, of feelings and of perceptual situations[[30]](#footnote-30). Furthermore, the distinction between a subordinate proposition and a gerundive complement has already been used to distinguish reports of semantic memory and reports of episodic memory correspondingly[[31]](#footnote-31). Thus, we can ask how Bernecker can refuse the semantic/episodic distinction but nonetheless introduce the fact/event distinction based on the same grammatical principle. Leaving these similarities aside, the division proposed by Bernecker is not free of problems:

1) First, the “grammatical” distinction between different kinds of memories and the characterization of them shows some inconsistencies. For example, it seems inappropriate to distinguish between non-conceptual memories (and therefore no-propositional) and conceptual ones taking as a starting point the grammatical criterion of the possible complements of the verb “to remember”. Maybe the choice of this grammatical criterion is the consequence of what Malcolm calls an overintellectualization or idealization of remembering, that is, the idea that a genuine or cognitive memory always requires a judgment (which would also explain the exclusive analysis of fact memory, the only propositional type of memory)[[32]](#footnote-32).Additionally, it is not clear neither that the proposed distinction employs a “grammatical” criterion nor that it indeed employs a single criterion: if the difference between fact and event memories is grounded on the syntactic use of a gerundive complement or a subordinate proposition (which besides does not seem to mark a real difference), the distinction between object and property memory does not seem to be syntactic. From an exclusively syntactic point of view, we can never remember a property if it is not a substantivated property; but the use of the English verb “to remember” seems to prove that we only remember a substantivated property of something and never only a substantivated property (like for example, “I remember the whiteness”). Therefore, the example given by Bernecker is incorrect: “a floppy ear” is just an object with a particular property (a noun with an adjective) and not a property per se, even if the whole nominal syntagm can be viewed, from a semantic point of view, as a property of Fido. So in the case mentioned by Bernecker, from a syntactic point of view, the subject remembers an object with a particular characteristic, and neither a property instantiated in a particular object (which would require a substantivated property) nor a property per se (which would be impossible on this approach). This leads us to think that the criterion for differentiating objects and properties is evidently from other order different from the grammatical one.

2) Even if we set aside these inconsistencies, it is not evident that a grammatical criterion would provide a good categorization of the different kinds of memories. Either it is symply a grammatical distinction of the possible complements of the English verb “to remember”, which seems not to be Bernecker’s purpose (and which would only have linguistic interest) or it has to assume an ontological distinction between properties, objects, facts and events, or at least a differentiation in the ontological kinds of which we can be aware. Nevertheless, Bernecker does not give any kind of justification for his approach[[33]](#footnote-33). For example, it is highly controversial from a metaphysical point of view wheter events supervene on objects or whether objects in fact supervene on events, whether there is a difference between events and facts, and whether an event is only a special kind of property[[34]](#footnote-34). Moreover, we can question the very division between facts, events, objects and properties. Besides these ontological problems, it is also questionable whether these four types correspond to different kinds of memories, or, for example, whether we could actually have a memory of a property without also having a memory of an object[[35]](#footnote-35). Due to the importance of this classification for the analysis, it would have been extremely useful for Bernecker to develop it in more detail, as well as to provide some grounding for it. In this sense, the semantic/episodic distinction comes with a longer tradition of analysis, not only at a conceptual level but also at an empirical one.

In conclusion, Bernecker not only does not give any arguments in favor of the proposed typology, but also does not develop an account for all of them, because he decides to focus his attention exclusively on fact memory. His justification is the following one: “Broadening the study to include object-, property-, and event memory would demand a discussion of the difference between conceptual and non-conceptual content, a much discussed topic of contemporary philosophy of mind. Yet since the conceptual/non-conceptual content distinction is not specific for memory, it would lead us away from our guiding question 'What is memory?'”.[[36]](#footnote-36) Nevertheless, as we saw, the problem is deeper than Bernecker’s suggestion, because it affects the very basis of the typology that he proposes.

For the moment, as we have seen, the object of his analysis is the occurrent use of remembering (and not its dispositional sense)[[37]](#footnote-37) and, within explicit and long term memory, fact memory. Within fact memory, Bernecker will only analyze:

- non-inferential memory and not inferential memory. This distinction, though not new[[38]](#footnote-38), can be questioned. First, the example he gives could be perfectly interpreted as a simple judgment made from information retrieved from memory: “I didn’t lock the door” could be simply the conclusion of reasoning from “I closed the door behind me and walked away”; and second, this distinction seems to assume on one side that there are basic or primitive propositions and other derived propositions, and, on the other hand, that only primitive propositions are actually stored in memory, as a singular item stored in a box.

- memory de re and not memory de dicto, where “the difference between memory de re and memory de dicto is that the former but not the latter allows for the substitution salva veritate of coreferential expressions”. Following Bernecker’s example, in memory “de dicto” we can remember that Marilyn Monroe is blond without remembering that Norma Jean Mortenson is blond. Memory “de re”, in contrast, is not sensitive to the choice of words because, as Bernecker explains, it indicates a connection between the subject who remembers and a fact in the world. However, here this might pose a problem, because fact memory is explicitly assumed to be propositional; in fact, in the later chapters, he mentions a list of principles which establish the type of deviation that is admissible between the initial proposition and the proposition that would correspond to the act of remembering. So, it is difficult to see how an analysis focused on propositions could be at the same time an analysis of memories de re and not of memories de dicto. For example, it is difficult to see in a propositional account of memory why my memories would not be sensitive to coreferential expressions: if once I learned that a woman called Marilyn Monroe was blond, how would my memory of this initial representation would be indifferently expressible in the following propositions: “I remember of Marilyn Monroe that she was blond”, “I remember of Norma Jean Mortenson that she was blond”, “I remember of the third child born to Gladys Pearl Baker that she was blond”, etc? The only way that I could remember the last two propositions in Bernecker’s account, if I did not learn at the same time that Marilyn Monroe was blond that Marilyn Monroe was Norman Jean Mortenson and was also the third child born to Gladys Pearl Baker, etc., would be in an inferential way from different memories (which besides would be in contradiction with Bernecker’s willingness to analyze cases of memory that are not inferential).

I have not yet even explained Bernecker’s proposal, but nevertheless I have already criticized almost all of his characterizations of the object of his analysis, with the exception of the broad categories of long term memory and explicit memory, certainly due to the fact that they are generally accepted and not questioned (but not for that exempt from any possible criticism). It is certain that any typology can be criticized, even the classical distinction between semantic memory and episodic memory; my intention here is not to make a strict critique of the typologies used by Bernecker, but simply to show that:

(a) the object of his analysis is based on a primary typology (objects, properties, events and facts) that is not as standard as the semantic/episodic classification and that seems more questionable, and thus requires require at least a brief justification, which unfortunately is absent from Bernecker’s account;

(b) the distinction between non inferential/inferential memory, first, rests on questionable assumptions and, second, could be better viewed, in Bernecker’s account, as a distinction between pure remembering and reasoning from memories than as a distinction between two different kinds of memory;

(c) the secondary characterization of the object of his analysis as memory de re seems to conflict with his grammatical account of fact memory. Even if this problem is not easy to solve, at least it would help to have a somewhat deeper characterization of memory de re vs. memory de dicto.

With these remarks out of the way, there is one remaining point that I want to develop before turning my attention to the causal relation and the content relation between the past experience and the memory of it. I refer to the distinction drawn by Bernecker between two different types of fact memory:

- reproductive memory: when the complement of the verb “to remember” is a simple fact, a reproduction of one of my previous thought contents, e.g., “I remember that Fido bit the mailman”

- meta-representational memory: when the complement of the verb “to remember” is not a fact but a previous representational state of mine, e.g., “I remember that I believed that Fido bit the mailman”. Metarepresentational memory makes available two kinds of information: a previous thought content which constitutes information about the world and also the attitude that I had towards that content which refers to one’s past mental state (in the example the attitude of believing). So, the object of a meta-representational memory is a propositional attitude[[39]](#footnote-39).

This distinction is not new; we can find it described roughly in Martin and Deutscher’s article, or in Broad’s book. Bernecker adds two remarks on this distinction:

- the first one is a fairly minor, but despite this, it threatens to conflict with other points defended by Bernecker. He points out that the language (that is, the propositional reports that we use to express our memories), is not always a reliable indicator “of whether someone occupies a state of meta-representational memory or reproductive memory.”[[40]](#footnote-40) At first glance, this statement seems to say something obvious: we cannot define a mental state simply by relying on a propositional report. However, the opposite criterion is precisely the criterion underlying Bernecker’s fundamental distinction in four types of memory. According to Bernecker language is a reliable source for distinguishing the different kinds of memory but not for distinguishing the two sub-kinds of fact memory, a conclusion which is problematic and somewhat contradictory.

- the other remark that Bernecker makes is that the distinction is only possible for some cognitive states, like thoughts, beliefs and inferences, but not for conative states, affective states and some other cognitive states like “doubting”, where they can only be recalled by means of meta-representational states. Here again the distinction is made at a linguistic level: “I cannot remember what it is that I regretted without stating that I regretted it. Likewise, whenever I recall what I doubted I have to state that I doubted it (…) I can entertain or express the belief that Fido bit the mailman by thinking or saying 'Fido bit the mailman' without mentioning the attitude.”[[41]](#footnote-41) For this last case, Bernecker tells us that we have the choice between merely reproducing the content and representing the previous mental state. But, does the fact of having a meta-representational memory rather than a reproductive memory depends entirely on our choice of including or omitting the words “I believed”? This is not very convincing. Besides, even at a linguistic level it is not certain that this possibility of choosing between a reproduction of the content or a representation of the mental state would be intrinsic to some cognitive states: we could easily imagine a situation where I cannot remember what it is that I believed without stating that I believed, for example when I do not any longer believe what I once believed.

Furthermore, one important point concerning this distinction is that it is not at all evident that meta-representational memory should be viewed as a kind of fact memory: even if one of its contents is a fact, the other principal one, namely the attitudinal content, suggests that meta-representational memory is a kind of event memory (in Bernecker’s typology). Furthermore, we could ask, following the difference gerundive complement / subordinate proposition establish by Bernecker, if these two propositions “I remember believing that Fido bit the mailman” and “I remember that I believed that Fido bit the mailman” would correspond to two different kinds of memory.

*A preliminary analysis of the conditions of memory*

For the moment, I only present the object of Bernecker’s analysis, namely fact memory, as well as the two subcategories of fact memory that he acknowledges: reproductive memory and meta-representational memory. For each of these subcategories he presents a preliminary analysis of the necessary (but not sufficient) conditions for a representation being a memory:

(A) For reproductive memory:

(1) S represented at t1 that p\*

(2) p and p\* are identical or sufficiently similar

(3) S’s representing at t1 that p\* is the cause of S’s representing at t2 that p

(4) p is true at t2

(5) p\* is true at t1

(B) For meta-representational memory:

(1) S represented at t1 that p\*

(2) p and p\* are identical or sufficiently similar

(6) S representing at t1 that p\* is the cause of S’s representing at t2 that he represented (at t1) that p.

Based on of this preliminary analysis it is not difficult to see that all of the main ideas of Bernecker ‘s work correspond exactly to with the main ideas of Martin and Deutscher’s article:

(a) two possible sources of remembering: facts in the external world and personal mental states, which corresponds, as we have already seen, to the most plausible interpretation of clause 2 of Martin and Deutscher’s account;

(b) a causal relationship between the past representation and the current one: clause 3 (Martin and Deutscher) and conditions (3) and (6) of Bernecker’s account;

(c) a similarity of content or structure between both representations: clause 3c (Martin and Deutscher) and condition (2) of Bernecker’s account;

(d) at least a true past representation: clause 1 of Martin and Deutscher and conditions (4) and (5) of Bernecker (who also requires that the past representation continue to be true at the moment of remembering).

Certainly there are some differences, and certainly these ideas are more thoroughly developed in Bernecker’s book (and due to the longer explanation they receive there, they are also more specified in more detail, which makes them more vulnerable to criticism). However, the fundamental schema is the same, as is the role played by the memory trace, as we will see next.

*The memory trace*

Just as the causal interpretation of memory is adopted because of its higher plausibility relative to the other possible interpretations (as we will see next), the idea of a memory trace is adopted because of the impossibility of explaining the memory causation without it. The argument is the following:

(1) There is no conclusive evidence in favor of memory traces, because they:

(a) has not been empirically established;

(b) cannot be established by introspection;

(c) are not necessarily implied by the concept of memory.

(2) However, the only other possibility of explaining memory causation, that is, the mnemic theory developed by Russell, is not satisfactory from a theoretical point of view; moreover, it collapses into the trace theory (as we will see later).

(3) In conclusion, the idea of a memory trace must be incorporated in a causal theory of memory.

I want to make two remarks concerning these points:

First, in Martin and Deutscher’s account, the idea of a memory trace is not only an implication of a causal model of memory but “an indispensable part of our idea of memory”. That implies the acceptance of “the existence of some sort of trace”[[42]](#footnote-42), even if, as we have already seen, it is not clear whether the trace refers only to a theoretical construct or also to a physical entity. Bernecker, however, denies that the existence of a memory trace, conceived either as a physical trace or as a conceptual entity, can be established from the concept of memory. The concept of memory only implies that “there is some form of causal link between the past and the present representation”.[[43]](#footnote-43) The idea of a memory trace is the best hypothesis concerning the nature of this causal link, but this does not mean that the concept of a trace is implicit in the concept of memory: “traces are a feature of the dominant hypothesis about memory, rather than a feature of the *concept* of memory”[[44]](#footnote-44).

Second, the mnemic theory, that is, the idea of a direct causality of the past experience in order to produce the memory of it, initially proposed by Russell, is ruled out not because of the traditional objection that a causal link over a temporal gap is impossible, as Martin and Deutscher do, following Broad’s critique[[45]](#footnote-45), but because of the impossibility of making sense of an alternative way of explaining the retention of information in order to account for the memory process. Bernecker points out that the only alternative that would in principle seem more viable it is the dispositional account of retention defended principally by Squire and Schumacher[[46]](#footnote-46). But the dispositional account, which says that remembering a past experience consists in exercising a disposition to represent that past experience acquired at the moment of the experience, supposes too that this capacity is retained from the moment of its acquisition, a retention that unfortunately cannot be explained without the introduction of something similar to a trace. This is why the dispositional account, and thus the mnemic theory, collapses into the trace theory, and this is why the trace theory is the only plausible explanation to account for a causal theory of memory.

But even if there are these minor differences between Martin and Deutscher’s and Bernecker’s accounts, the central point is the same: a causal theory of memory is only feasible given that we accept the notion of a memory trace. Not only is this idea similar, but so are the two principal functions that traces accomplish: the propagation of information (that is, the conservation of the content) and the production of states of recall (that is, the causal bond). However, Bernecker does not attribute both functions to the same notion of the memory trace: in order to produce states of recall, traces have to be physical, like structural modifications of synapses, but in order to communicate information, they have to be mental states. So this division implies the acceptance by Bernecker of a dual nature for traces: the trace has a physical aspect and at the same time a mental aspect, and each aspect has a different function. First, we could ask why he doesn’t conceive that a physical trace could communicate information; and clearly the answer has to do with a conception of the information as semantic content[[47]](#footnote-47), which need not be conceptual for all the types of memory, but only for fact memory. This is why Bernecker distinguishes two types of traces: dispositional beliefs for the traces of fact memory and subdoxastic states for traces of object-, property-, and event memory. Second, it’s the differentiation between the level of physical implementation or realization and the computational level, for which he clearly adopts a semantic approach (and not a connectionist one), that in a certain sense leads him to distinguish these two “aspects” of the trace and to attribute to each a specific function. I digress briefly to account for this division in different levels of explanation.

Bernecker, as do most authors (for example D. Marr and Z. Pylyshyn[[48]](#footnote-48)), distinguishes three levels of analysis of the mind: for an intentional process “first, there is the intentional level of everyday psychology, in which we talk about people’s beliefs, memories and other such intentional states. Second, there is the computational level which explains how intentional states are realized by means of computational operations. Finally there is the level of physical implementation or realization.”[[49]](#footnote-49) However, as I have already pointed out (and as I will discuss further below), Bernecker doesn’t maintain his explanations at one level of analysis but quite the opposite: he first separates the object of analysis at different levels but subsequently mixes these different analyses up.

Returning to the computational level, he distinguishes two possible accounts: the symbolic and the connectionist approach. The first approach holds that mental representations are symbolic structures that have semantically evaluable constituents and that mental processes are rule-governed manipulations of them. In this case, Bernecker considers that the format of the mental events stored in traces are not finished sentences or propositions but rather fragments of these complex syntactic structures. The second approach, maintains that mental representations are realized by patterns of activation in a network of model neurons and that mental processes consist of the spreading activation of such patterns. Bernecker points out that there are many arguments in favor of connectionism, including the ease with which it allows an explanation of the reconstructive nature of memory, because retrieval is viewed as a reconstruction of information rather than the location of a discrete stored item. Nevertheless, even if he assumes that one fundamental characteristic of memory is its reconstructive nature, he doesn’t adopt a connectionist approach but rather a symbolic one. He holds that the connectionism is incorrect because it assumes that memory always reconstructs previous recorded information, whereas sometimes memory effectively reproduces it. In his own words: “Constructivists are right on the mark when they claim that the retrieved information does not have to be identical with the encoded information and that information processing is an essential aspect of remembering. However, constructivists seem to forget that the factivity constraint on memory demands that the memory output be true and sufficient similar to the input.”[[50]](#footnote-50) However, as we will see below, the construction of the memory event is not an exclusive property of connectionism (and it is not necessarily opposed to the reproduction of a past event); a symbolic model is also perfectly compatible with it, and therefore this minor criticism of connectionism, and thus the “choice” of the symbolic model, are not well grounded.

Returning to the differentiation of levels mentioned above, this allows us better to understand why Bernecker separates the two functions generally attributed jointly to the trace: each function corresponds to a different level of analysis of the trace. There is, however, a problem: if these two functions are separated, what is the bond that allows the preservation of content if, on the one hand, the physical “aspect” of the trace exerts a causal role but doesn’t propagate information, and, on the other hand, the computational “aspect” of the trace doesn’t have causal power? How could the trace conserve its unity, which is indispensable for explaining the unity of the memory event? Dividing the trace into “different aspects” which accomplish different functions without specifying anything more about it only generates more confusion concerning not only the status of the trace but also the nature of the causal bond between the past event, the trace and the memory.

At least, it seems clear that the trace, in each of its aspects, is more a vehicle or a medium for maintaining the causal link and propagating information than an autonomous entity that starts another causal connection independent from the one which gives rise to it. This idea is clearly deduced not only from the two roles attributed to the trace, but also from the idea that the trace has to be introduced into the causal theory of memory in order to explain the causal link between the initial representation and the current representation (which will be the principal causal link). This conception of traces is exactly the same as in Martin and Deutscher’s account.

After considering the notion of the trace in Bernecker’s account, I will develop, as I did for Martin and Deutscher’s, the two relevant aspects: causality and similarity of content.

*Arguments in favor of a causal theory*

I begin with the first aspect. In relation to the causal theory of memory, Bernecker distinguishes two arguments in favor of it and three arguments against it. The two arguments in favor of the theory are, on the one hand, the argument from counterfactuals that he attributes to Martin and Deutscher (incorrectly as I have already noted), and on the other hand, an argument from personal identity given in Shoemaker’s paper “Persons and their past”[[51]](#footnote-51). As far as arguments against are concerned, he finds two in Malcolm’s article “A definition of factual memory”[[52]](#footnote-52) and another in Squire’s paper “Memory unchained”[[53]](#footnote-53). The five arguments, however, are dismissed because none of them, according to Bernecker, succeeds in confirming or refuting the causal theory of memory. The two objections made by Malcolm concern the causal relation between the past and the present representation: the first is that this relation is logical and not metaphysical, and therefore not causal; the second, is that causality has a nomological character that is absent from causal claims about memory. Both criticisms are, however, rejected by Bernecker. The third criticism comes from the dispositional account but, as we saw before, this can’t be explained with a trace-free notion; it is also the case that in this account the difference between temporary and permanent forgetting becomes meaningless. As far as arguments in favor are concerned, we have already mentioned that attributed to Martin and Deutscher; as far as Shoemaker’s argument is concerned, in general terms, he considers that only if we understand the notion of memory as implying a causal process can we account for the role played by the notion of memory in judgments of personal identity. However, Bernecker points out that this explanation is circular, because we can’t appeal to autobiographical memory in explaining our numerical identity through time due to the fact that autobiographical memory already presupposes personal identity. Nevertheless, it is unfortunate that Bernecker only considers philosophical arguments in favor of the causal theory without taking into account any empirical considerations.

Setting this shortcoming aside and returning to the general strategy employed by Bernecker to justify the adoption of the causal theory of memory, we can reformulate his argument as follows:

(a) There is not currently any argument in the literature that confirms the causal theory of memory;

(b) But neither is there an argument that refutes the causal theory;

(c) The arguments in favor of alternatives accounts of memory retention are flawed;

(d) So the causal theory of memory is the theory more plausible.

The point (c) refers to two versions of the epistemic theory of memory, according to which memory is “retained (or long-standing) knowledge, that is, knowledge which was previously acquired and preserved”[[54]](#footnote-54). The first one is the pure retention condition, which can be formulated in these three forms:

(1) At t2 S has not forgotten that p.

(2) There is no time t1,5, between t1 and t2, when S did not know that p.

(3) At t2 S believes that at t1 he knew that p.

The first and the second are rejected because it’s perfectly possible to conceive of a case where a person remembers that p but he has forgotten it in the interval between the prior and the present knowing. Another reason for rejecting to the first point is the circularity of the argument, because the notion of memory is defined by appealing to the notion of remembering. The third argument seems to be or too stringent or too liberal, in the sense that someone can remember that *p* without believing that he previously knew that *p* or, on the other hand, he could not be remembering that *p* but believing that at t1 he knew that *p*. Even if the last two counterarguments don’t seem very convincing, the general criticism that Bernecker gives of the pure retention condition is that it can’t explain how knowledge is retained, which is exactly the same criticism that he gives of the mnemic theory of memory (which express more or less the same idea). As this version of the epistemic theory of memory is formulated, this last remark is convincing.

The second version of this theory is the evidential retention condition, which tries to explain the idea of sameness of knowledge in terms of sameness of reasons or justification. The formulation is:

(4) S’s grounds for knowing at t2 that p are identical to S’s grounds for knowing that p at t1.

The problem is not only that the grounds can be enlarged, but also once again that the conditions can be satisfied without the subject remembering that *p*, simply because he forgot *p* between the two intervals. And even if these two problems can be avoid with a reformulation of the evidential retention condition, as some authors attempted to do[[55]](#footnote-55), neither of these reformulations can avoid the case where we can lose some or all our justifications for believing that *p* and so, come to be incapable of knowing that *p* despite continuing to remember that *p*. Bernecker concludes that the evidential retention condition fails because it presupposes that memory implies justification, while he maintains precisely that this is not the case.

The conclusion of this indirect argument is that the only possible explanation of remembering is the causal account, which is why Bernecker adopts it. In order to analyze the causal bond, he decides to assume a notion of causality as a necessary and sufficient condition, after ruling out the possibility of understanding causality in terms of:

- only a sufficient condition;

- only a necessary condition;

- counterfactuals;

- probabilistic analysis.

In order to analyze more specifically this notion of causality, he takes from psychology the classical distinction between three types of retrieval:

- free recall: where the subject tries to remember information without any external information to help him;

- cued recall: when some explicit information which is not identical to the stored information is given to the subject to help him to remember;

- recognition: when the very piece of information the subject has previously learned is presented to the subject in order to ask him whether it is the piece he is trying to remember.

We can nevertheless question this typology: the cue need be neither external (it might be another thought, for example), nor explicit to the subject (who might perfectly not be conscious either of the cue or of the link between it and the memory). I will however not try to settle this question here.

For cases where a piece of information (either identical or not) is required, Bernecker differentiates four ways in which retrieval cues and memory traces can cooperate in the formation of what he calls “states of recounting”. Two quick explanations before developing these cases: first, Bernecker introduces the word “recounting” in order to refer to “a person’s activity of reporting past events which leaves open whether the report qualifies as genuine memory or whether it is nothing but the repeating back of a previously presented retrieval cue.”[[56]](#footnote-56) Second, for Bernecker, unlike for Martin and Deutscher, a complete or strict prompting[[57]](#footnote-57) before a subject can recount certain event doesn’t necessary mean that the subject is not remembering, as we’ll see next.

The four possible interactions between cues and traces are the following:

(1) causal overdetermination: when the memory traces and the retrieval cues are independently sufficient and simultaneous causes of recounting. In order to give an example, Bernecker takes up again Martin and Deutscher’s thought experiment about a person who forgets an experience from his past due to an accident but whose friend, to whom he had related this experience before the accident, describes it to him. As we can see, this kind of example based on past experience (or what Bernecker would call “events”), is absolutely out of place, because the object of his analysis was defined exclusively as facts. In any case, we can reformulate the example and suppose that what the subject Oscar forgets due to his accident is a simple fact about the world. On a later occasion, a friend tells this fact to Oscar who doesn’t remember it yet, and at the very moment when his friend is about to retell the story of the accident, Oscar’s memory trace about this fact revives. For Bernecker, the two causes are independent of each other and, because they are simultaneous, they are both causes of Oscar’s recounting the fact (even if he can’t add any more detail than what his friend told him).

(2) preemptive causation: when the memory traces and the retrieval cues are independently sufficient but successive causes of the recounting. In this case, the recounting counts as a genuine instance of remembering only when the memory trace is reactivated before the cues exert their causal influence.

(3) causal supplementation: when the memory traces and the cues are sufficient to produce a state of recounting only jointly but they do not interact before bringing about their effect. The example is the same as in the first case, because the two causes are independent and simultaneous; the only difference is that an independent contribution from each is required to produce the state of recounting.

(4) causal reinforcement: when the memory traces and the cues are sufficient to produce a state of recounting only jointly, but they need to interact before bringing about their effect. For example, when a weak memory trace is strengthened by a weak retrieval cue.

Concerning this classification, I must point out that the debates on overdetermination and causal preemption are far from resulting in a consensus. Loeb[[58]](#footnote-58) had already argued for distinguishing, in any causal theory between causes and causal overdeterminants. But even if we accept that overdetermination is genuine causation, there is also a disagreement about the individual or collective nature of it: those who state that each overdeterminant causes the effect individually would deny (3) and (4); those who assert that the effect is always the result of the collective causation would deny (1)[[59]](#footnote-59). To all these conceptual debates we could also add the impossibility of applying these logical categorizations to particular cases of memory, which calls their utility into question. For example, the exemplification of (3) is not very convincing: “Suppose that Bert’s starting to tell the story of the accident and the revival of Oscar’s memory trace take place simultaneously but independently. The two causal chains do not interact but they jointly bring about their effect, namely Oscar’s recounting of the story of the accident.”[[60]](#footnote-60) Even if it is a logical possibility, how can we really be sure that in the case of memory is not the cue that helps the trace to bring about the memory? In fact, we could even ask if a single cue could be the sufficient and independent cause of recounting. Thus, this classification, even if interesting and original, might be controversial.

Anyway, the conclusion at which Bernecker arrives is the following: “for a memory trace to give rise to genuine remembering it must be an independently sufficient condition (which is not preempted by another independently sufficient condition) of a state of recounting or at least a necessary condition of such an independently sufficient condition.”[[61]](#footnote-61) Clearly, the first case refers to ways (1) and (2), when the memory trace is a sufficient and independent condition simultaneous or previous to the retrieval cues; the second case refers to ways (3) and (4), when between the conditions that effectively produce the state of recounting the memory trace is a necessary one, either because its independent contribution is necessary or because it acts jointly with the retrieval cues. The description of this last case is nonetheless a kind of reformulation of Mackie’s definition of causation: a cause is always at least an insufficient but necessary part of an unnecessary but sufficient condition for some event. This definition (which Mackie refers to as an “inus condition”) is, however dismissed by Bernecker, who adopts a different version of this definition, which does not correspond to Mackie’s formulation[[62]](#footnote-62).

In short, I conclude that, on one hand, this classification is interesting because it’s the first time that an account of the causal theory of memory conceives of different possible types of interactions between the memory trace and the retrieval cues in producing the memory event. On the other hand, the explanation follows the path outlined by the traditional causal theory of memory, because the memory trace continues to be the key element that finally determines when a state of recounting is a genuine act of remembering. We might formulate this point as follows:

For a state of recounting to be a genuine act of remembering it is a necessary condition that the corresponding mental trace plays a causal role in producing that state of recounting. This causal role consists of being:

- either an independently sufficient condition (which is not preempted by another independently sufficient condition)

- or a necessary condition of such an independently sufficient condition

As we have already mentioned, this centrality given to the trace is nothing more than the consequence of the centrality given to the past representation in the causation of the memory event, in conjunction with the adoption of the trace theory as the only plausible interpretation of the causal theory of memory.

I turn next to the other central point of interest: similarity of content. But once again I must briefly postpone the analysis in order to make some remarks on two points: one concerning the object of the memory; the other concerning the truth conditions of reproductive memory and meta-representational memory.

*The object of memory*

In relation to the this point, Bernecker criticizes representative realism about memory, which holds that “one remembers something not by way of being directly aware of that thing, but rather a mediated image which represents that thing.”[[63]](#footnote-63) In analogy with the sense data theory of perception, this memory-image is the memory-datum which is brought about by the memory trace. According to Bernecker, the representative realism would hold that the memory-image of something x:

(i) is a mental entity which serves as the object of a direct memory;

(ii) shares numerous properties with some prior perception of x;

(iii) is causally linked with some prior perception of x.

Martin and Deutscher would fit in this category, with the addition that they maintain that the memory-datum and the memory-trace not only share numerous properties but are structural analogues of the original experience, an idea that Bernecker considers “as implausible as claiming that a virus carrying a contagious disease from one patient to another must be structurally analogous to the symptoms of the illness it transmits.”[[64]](#footnote-64). Bernecker, in contrast, even if he considers that these memory-data are necessary for remembering and that they also determine “the way the thing appear to us”, denies that they are the primary objects of our memories, maintaining that they are just the vehicles of the remembered information: “I am aware of the past event by internally representing the event, not by being aware of the internal representation of the event.”[[65]](#footnote-65) However, as Bernecker explains, one problem with direct realism is that this theory is only characterized in negative terms, that is, by denying the explanation offered by the representative realism, as Malcolm earlier pointed out[[66]](#footnote-66). Even if he defends direct realism from the objections generally made to it, he never gives an explanation of how is it possible to be directly aware of a past event in the act of remembering, as happens on all direct realist accounts.

I will not say say much more about the direct realism problem because it’s a complex debate that is beyond the scope of this work. The only thing that I want to point out introducing is that Bernecker, due to the lack of a positive characterization of the direct awareness of the past event, does not seem to say anything more than that objects of our memories are not internal representations of the past event but are rather past events themselves. Even if he pleads for an eternalist view, that is, the idea that events do not cease to exist when they cease to be present or when there ceases to be evidence for them, that, once an event occurs, it exists eternally, to argue for the existence of the past experience seems incompatible with the notion and role of the memory trace. Let me explain why. In response to a criticism made by Robinson (1994)[[67]](#footnote-67), who maintains that “I cannot make any sense of the idea that a past object could constitute the content of a current mental state without the aid of a current form of representation”, Bernecker declares that “Robinson’s worry is groundless, for direct realism is not committed to denying that cognition involves 'a current form of representation'. But according to direct realism these representations are the vehicles of perception [and memory][[68]](#footnote-68) rather than its objects.” He continues: “And to say that perception directly acquaints us with a past object is not to say that it provides some inexplicable immediate awareness or experience of the past object. We are directly aware of the past object in the sense of not having to rely on present evidence”[[69]](#footnote-69). Even if it’s clear that with the words “present evidence” Bernecker alludes to the classical idea of a memory of something as an immaterial entity housed in the mind (like an image, a copy, a representation), it’s nevertheless odd that he denies the existence of “present evidence” even while maintaining a notion of memory traces: is the memory trace not a sort of “present evidence” from which the past event is inferred in an unconscious way? And if not, what is the sense of introducing the notion of a “memory trace”? And, moreover, what else could it mean to say that the trace accomplishes the function of propagation of information?

This suggests that it’s not possible to introduce the notion of memory trace along with an eternalist view about the existence of the past events. A direct realist account would only be compatible with a trace theory if we interpret the idea of the past experience being the direct object of memory in a logical sense, or in a grammatical sense (which, besides, would agree with his grammatical approach to memory). If so, the difference between representationalism and this direct realism could only refer to the logical or grammatical opposition between “S remembers the past event x” and “S remembers the representation of the past event x”. But once more, this difference seems to blur, because both approaches can jointly state that the logical object of the memory is the past event, being this past event a representation from an ontological point of view. Why? Because if Bernecker accepts the idea of memory traces as vehicles of information, he is forced to admit that:

(a) the past event could only be the object of a memory from the information propagated by the corresponding memory trace;

(b) once the past event becomes the object of a memory, it cannot have an ontological status different from a representation.

That is why it is difficult to see how the combination of this kind of “direct realism” with a theory of trace could be differentiated from what is generally understood as a representative realism, such as Martin and Deutscher’s account. Taking up his characterization of the “representative realism” enumerated before:

- (i) would be the only point of difference with his account; however:

(a) it’s not necessary that in a representationalist account the object of a memory must be a mental entity or image that represents the past event that is housed in the mind; even if this kind of representationalism was very common (especially before the 1960s), other representationalist accounts consider that the past event is the object of a memory in the logical or linguistic sense proposed by Bernecker, this past event being the result of an unconscious process of representation from the information propagated by the memory trace. Even if Martin and Deutscher don’t specify these points, this explanation of the representative realism would be absolutely compatible with the account outlined in their article.

(b) this account of representative realism will be exactly the same as Bernecker’s account, because, as we have seen, there isn’t any other way of interpreting the conjunction between the special “direct realism” badly defended by Bernecker and the trace theory of memory.

- The requirement (ii) of similarity or identity between the two representations (the perception and the memory) will be exactly the same as the condition (2) required by Bernecker’s account: “p and p\* are identical or sufficiently similar”; the only difference would be that for Bernecker *p* is a proposition, that is, a state of recounting, whereas in a representative account this is not necessary. In fact, taking this point into consideration, we reaffirm the idea described above: if the object of a memory is a past event and this past event can only be described in a proposition, a direct realist approach could only mean that the past event is nothing more than the logical object of the memory.

- Here again point (iii) is similar to conditions (3) and (6) required by Bernecker (one for reproductive memory, the other for meta-representational memory). The only difference with Bernecker, at least with respect to Martin and Deutscher’s representationalism, is that he denies that the memory trace is similar in content to the past event; the memory trace would be only a vehicle that allows the propagation of information. However, I precisely tried to show that for Martin and Deutscher the trace is also a simple medium that allows the conservation of the structure of the past experience in order to transmit it to the structure of the memory event. In this sense, the trace is also a kind of vehicle of information. The problem is that even if Bernecker finds the idea of memory-traces or memory-data as structural analogues of the past experience totally implausible, it’s evident that in his own account he has to admit at least a sort of correspondence between this information propagated through the memory trace and the content of the past event and the memory of it. If not, we could ask once more what the idea of a trace that accomplishes the function of propagation of information could mean. Furthermore, we could ask, as we will see next, if the symbolic approach chosen by Bernecker at the computational level (the level that would correspond to the mental aspect of the trace whose function is conceived of as the propagation of information) and the notion of trace that it implicates, as fragments of complex syntactic structures like sentences or propositions[[70]](#footnote-70), would really allow him to refuse the idea of a trace with a semantic or a syntactic content similar to the past fact.

*Truth conditions of memory*

The other topic that I want to discuss is the truth conditions of the memory. Bernecker points out that these vary depending of the kind of fact-memory at issue[[71]](#footnote-71). For the case of reproductive memory, “apart from the truth of the content, the truth of a reproductive memory claim depends on whether one represented the content in the past.”[[72]](#footnote-72) So, the truthmakers of the reproductive memory are the following ones:

(a) a previous representation of the factual content;

(b) the truth of the content;

(c) a similarity between the content of the past and present thought-tokens.

For the meta-representational memory:

(a) a previous attitude (like believing, doubting, etc) towards a factual content;

(b) the truth of the factual content;

(c) a similarity between the first-order content of the past and present thought-tokens;

(d) a similarity between the original attitude and the attitudinal component of the present representation.[[73]](#footnote-73)

*Similarity of content*

Now we can return to our principal topic: identity or similarity of content. Bernecker’s account is differentiated from what he calls the xerox model, which has two versions: for the first one, the persistence model, recall is a revitalization of an existing trace; for the second one, the reproduction model, recall is constructed as the reproduction of the original representation. But for both submodels:

(a) memory is viewed as a passive process for registering and storing items of information;

(b) discrete items are used as memory input, allowing a quantitative approach to memory, which leads to measuring memory effectiveness according to the number of recovered elements and conceiving of forgetting as information loss.

So Bernecker’s position is different from connectionism, which is presented as a radical constructivism that would allow the modification and generation of content, and also from the xerox model, which, in contrast, does not allow any kind of change into content. The general principle observed by Bernecker is that the content of the current token cannot increase or enrich the content of the past token. It allows only:

(a) the loss or omission of information;

(b) some moderate transformations that involve neither a loss nor an increase of information;

(c) the condensation of information.

These modifications of content are due to the different ways in which memory edits encoded information. Bernecker center his attention on three of these:

(A) Cognitive dynamics: These mechanisms allow varying the characters of our judgments in order to express the same content. The variations refer to specific changes in the articles, in the time tense, etc (as we’ll see later).

(B) Condensation: The condensation enables memory to refer not only to a particular event that happened at one time but to periods of time. In these cases, the content of the condensed memory has no single counterpart in the past, but it is the result of a kind of conjunction and inference. For example, a genuine memory like “I remember having disliked eating vegetables as a child”[[74]](#footnote-74) doesn’t necessary imply that in my childhood I believe that I didn’t like vegetables, but only that I thought at least that “I don’t like Brussel sprouts”, “These endives are too bitter”, “Carrots are too sweet”, “That spinach has a disgusting texture”, etc. As we can see, this seems a key point because it clearly differentiates Bernecker’s account from Martin and Deutscher’s. In Martin and Deutscher’s article, it was clear that there was an univocal relation between a past event and the memory of it; Bernecker on the contrary seems to deny this one-to-one correspondence due to admitting the possibility of condensation of information. However, we could perfectly reconsider this mode of editing information not as a memory processing but as a simple mode of inferential reasoning from different memories traces (or what Bernecker insists on calling “inferential memory”, which is not the object of his analysis). This is the same point I highlighted when discussing the difference between inferential and non-inferential memories. Nonetheless, the idea of condensation, either as a memory mode of editing or as a simple mode of inferential reasoning, gives rise to a problem for Bernecker’s account: once we arrive at a memory produced by a process of condensation, like the example “I remember having disliked eating vegetables as a child”, does this memory trace produce a new trace, which would be considered as the cause of a subsequent recall of this condensed memory? Bernecker does not provide an answer to this question. This is perhaps because, on one hand, he does not analyze memory from a dynamic point of view; and because, on the other hand, it seems that his only interest in adding this mode of editing consists of explaining the possibility of remembering universal statements from particular ones, allowing also in the sentences the change from particular concepts to universal concepts.

(C) Schematic processing: Bernecker considers this to be the most invasive mechanism among the three; it refers to the capacity of our memory to edit information in accord with cultural schemas. It’s not only that we tend to pay attention to and register what fits more closely with our personal and cultural point of view; the principal point is that, as time goes by, the omissions and inaccuracies are systematic and corresponds to the filters and patterns imposed by our personal and cultural schemas. In these cases, quoting an example taken from Bartlett, Bernecker seems to admit that remembering is a reconstructive and generative process, which implies a constant effort of meaning and imagination. Nevertheless, with the exception of the allusion to the general principle of omission, Bernecker doesn’t develop any principle that would correspond to the schematic processing, as we’ll see later. That’s why we could ask to what extent this “mode of editing” is really compatible with his account, especially given that schematic processing allows the increase and transformation of information, which is not, however, admitted by Bernecker. For example, citing the experiment performed by Bartlett, which consists of having some subjects recall a story that fit in poorly with their cultural background, he notes that “ 'something black came from his mouth' in the original story becomes 'he foamed at the mouth' or, in some stories, 'he vomited'. […] Furthermore, some subjects introduced inferences in an attempt to explain incongruous features of the story. For example, when one of the men declines to go off to battle, he states 'I will not go along. I might be killed. My relatives do not know where I have gone. But you -he said, turning to the other-, may go with them'. One subject recalled this last passage as 'But you have no one to expect you' and another subject recalled it as 'You have no parents'”[[75]](#footnote-75) As we’ll see later, none of these changes can be neither explained by a principle of semantic similarity nor admitted as a case of genuine remembering on Bernecker’s account[[76]](#footnote-76).

Bernecker invokes two arguments to show that genuine remembering involves condensing and editing information:

(1) Narrative self-definition: To form a coherent autobiographical story, memories have to be condensed and filtered through our cultural schemas, and it’s not necessary that each memory be an exact reproduction of the corresponding past experience or event. As Bernecker says, “life-stories stress the relevant aspects of one’s past experiences, depict behavioral patterns, and identify representative judgments and characteristic preferences. In the process of editing, anomalous events may be omitted, typical ones underlined, and other changes undertaken to increase the smoothness and intelligibility of the narrative.”[[77]](#footnote-77) I must nevertheless make two points: first, this argument once more applies only to events and experiences, and not to the kind of memory on which Bernecker focuses, that is, fact memories. Second, none of the principles of semantic similarity that I will enunciate below permits all the changes that he seems to admit here, neither the fact of refusing the increase of content.

(2) Clutter avoidance: If we consider the overwhelming amount of information we receive every day and the limit not only of our capacity for storing information but also for retrieving it, it is indispensable to condense and compress the incoming information.

Finally, I quote the different principles of semantic similarity that Bernecker develops in order to investigate the extent to which two content tokens may differ from one another while one of them still counts as a memory of the other. The principles are the following:

(1) Principle of tense adjustment. This refers to changes of time tense: memory may replace (i) statements in the (continuous) present tense with statements in the (continuous) present perfect; (ii) statements in the present perfect with statements in the past perfect; (iii) statements in the future tense with statements for the present tense.

(2) Principle of article substitution. (i) demonstrative and possessive pronouns may be replaced by indefinite and/or definite articles; (ii) definite articles may be replaced by indefinite articles, and (iii) definite and indefinite articles may be omitted altogether.

(these two principles are obligatory)

(3) Principle of universal instantiation: statements about individuals (F(k)) may be substituted for universally quantified statements (∀x F(x)).

(4) Principle of existential generalization: existentially qualified statements ((∃ x F(x)) may be substituted for statements about individuals ((F(k)).

(5) Principle of conjunction elimination: conjunctive statements (p ∧ q) may be replaced by statements mentioning only one of the conjuncts (p or q).

(6) Principle of conceptual universalization: up to a point, particular concepts may be replaced by implied universal concepts.

(7) Principle of omission: our memory may omit contextually irrelevant information. However, the amount and the type of information that can be lost in the course of remembering is relative to the context of remembering.

(8) Principle of negation elimination: a doubly negated statement (~~p) may be replaced by a positive statement (p).

(9) Principle of contrastive stress: diachronic thought-content-tokens which differ only with respect to their contrastive stress are sufficiently similar to fulfill the truth.

(10) Principle of synonym substitution: memory may replace terms with their known synonyms without thereby violating the truth condition of memory.

Principles (1) and (2) have a special status, because their use is obligatory in most circumstances. They govern mode of editing (A) (cognitive dynamics). Most of the other principles can be considered as principles applied in condensation. We can also see that they are principles that govern propositions and not narratives. Furthermore, none of them corresponds to schematic processing, and none of them needs support in terms of the notion of narrative self-definition. This leads us to conclude that these two ideas (which besides are based on the idea of increase or change of the content) are invoked in vain by Bernecker: they can neither be admitted in his account nor used as a foundation of some of its elements[[78]](#footnote-78).

Concerning the attitudinal component, Bernecker points out that in some contexts, the similarity between the original attitude-type and the one identified in a meta-representational memory is also sufficient. This similarity will depend on the classification of attitudes we take into consideration. We can divide attitudes into cognitive attitudes (which have to match an independently existing world) and conative attitudes (which bring about changes in the world), and at the same time we can distinguish between simple and complex attitudes. Another kind of classification could be centered on the functional roles accomplish by the attitudes, which are characterized by their inputs and outputs. For example, if we choose the first characterization, “I wanted vanilla ice cream” will be a genuine memory of “I decided in favor of vanilla ice cream instead of chocolate”, because they are both conative attitudes and the only difference lies in the simplicity/complexity of each attitude (deciding being a complex form of wanting).

Even if they are applicable to most contexts of remembering, there are some contexts in which they don’t apply. For Bernecker, the validity of these principles, that is, the extent to which the past and the present content tokens or attitudes need to be identical or sufficiently similar in order to constitute a genuine memory, are context-dependent. Bernecker defends what he calls a memorial contextualism or a pragmatic sensitivity account of memory. We can enumerate his principal ideas as follows:

- the truth conditions of an utterance or a thought are context-insensitive;

- the determination of which of two disjuncts of the content and the attitude conditions is meant in a particular memory ascription is context-sensitive;

- context is defined as conversational context, which includes both the subject context, that is the context of the rememberer himself, and the attributor context, that is, the context of the memory ascriber.

- the conversational context is defined in terms of the intentions and interests of the rememberer as well as those of the bystanders: for example, if the agent’s intention is to recount something by heart, like a poem, the standards for fulfilling the truth conditions of memory are higher than if he merely tries to give a rough account of what he has learned. As well, the intentions and interests of the bystanders affect the standards that the remembered has to reach: if the content of a memory is too unspecific to be of any use, the bystanders have a right to deny that the rememberer is actually remembering (that is, even if he’s actually remembering something, he doesn’t remember what he “has to” remember in that specific context).

Consequently, the concept of memory is held to be relationally absolute: it’s absolute because remembering that *p* implies accurately representing that *p*; but it’s also a context-dependent concept, because what counts as an accurate or veridical representation of p is defined to some degree relative to the interests and intentions of the rememberer and the attributors of the conversational context.

*Some concluding remarks*

The notion of memorial contextualism is the most interesting innovation that Bernecker introduces in his account and in the philosophy of memory in general, maybe because it’s one of the scarce attempts to introduce a relativist element in a field that has almost always been reluctant to accept any sort of relativism. Nonetheless, even if Bernecker’s account can be considered to be innovative in some sense (and not only due to this last point), the basic structure underlying his theory does not differ from that proposed by Martin and Deutscher, as I have already shown:

(1) A true representation of the fact at both moments: at the first moment of representing it and at the moment of remembering it;

(2) A causal relation between the past representation of the fact and the memory of it generated by a memory trace, where the memory trace has to be an independently sufficient condition (which is not preempted by another independently sufficient condition) of a state of recounting it or at least a necessary condition of such an independently sufficient condition;

(3) A similarity between the content of the past representation of the fact and the memory of it, a similarity that:

a- is inconsistent with enrichment of content;

b- is governed by some specific principles whose application is sensitive to the conversational context.

(4) A biunivocal correspondence between a past representation and the memory trace of it but a univocal correspondence neither between the past representation and the memory nor between a trace and a memory (this, because Bernecker does not want to rule out the possibility of a trace that combines with other traces to give rise to a condensed memory, even if he doesn’t specify this);

(5) The conception of a memory trace as a single and independent unit that survives unchanged over time;

Throughout the analysis of Bernecker’s proposal, I’ve pointed out it’s similarities with Martin and Deutscher’s account, even while stressing the differences between the two theories. There is no need to repeat either the internal inconsistencies or the unstated assumptions of Bernecker’s theory here; I only want to point out that:

- as I have already mentioned, Bernecker’s theory suffers from what Malcolm calls an “overintellectualization” of memory, that is, the idea that memory is essentially propositional: “Philosophers fell tempted to concentrate in a portion of the total range [of the uses of 'memory' and 'remember'] and fix on it as 'memory in the fullest sense' or as 'genuine' memory. But these stipulations are injustified”[[79]](#footnote-79);

- many of the arguments and examples employed by Bernecker don’t correspond to the object of his analysis, that is, memory of facts, because they refer to memory of events, notions that were clearly distinguished from each other in the first part of the book;

- even if the analysis is done at a computational level and from a symbolic approach, Bernecker seems to have problems staying at a single level of analysis, as I have already noted. Besides, the grammatical criterion chosen to differentiate the kinds of memory limits him from the beginning in a way that he clearly doesn’t always respect.

- owing to this mixture of levels of analysis, the notion of a trace is dissociated in two levels, a dissociation that doesn’t help to clarify the status of the trace, the propagation of information through it, or its causal role. To this, we have to add the criticism that he views the trace as having a content similar to that of the representations, while also declaring paradoxically that traces are fragments of complex syntactic structures like sentences or propositions[[80]](#footnote-80).

**Going a little (but not too much) beyond the traditional causal theory of memory**

As we have seen, in Martin and Deutscher’s account, as well as in Bernecker’s, the content or structure of a memory has to be identical or similar to that of the original representation, but it can neither be enriched nor undergo any other sort of transformation. This is because the central element, considered not only as the principal cause but also as the determining factor of the content of the memory, is the past event[[81]](#footnote-81); the trace is only a vehicle for the causality and content of the latter, even if it is the proximate efficient cause of the memory.

Other recent versions of the causal theory of memory try to move away from the idea of a simple reproduction of content with the aim of reconciling the old assumptions of the causal theory with the possibility of a transformation and enrichment of content. The causes of these reconsiderations are multiple: the influence, on one hand, of empirical studies and theories in psychology; on the other hand, of Lackey’s article[[82]](#footnote-82) on the philosophy of memory, which defends an idea of memory as a generative source not only of the epistemic status of a belief but also of beliefs themselves; and finally the questioning of a model of memory that seems very far from the real operation of memory. We might justly say that, on Martin and Deutscher’s theory, any remembered would be like the character of Funes el Memorioso:

*Éste, no lo olvidemos, era casi incapaz de ideas generales, platónicas. No sólo le costaba comprender que el símbolo genérico perro abarcara tantos individuos dispares de diversos tamaños y diversa forma; le molestaba que el perro de las tres y catorce (visto de perfil) tuviera el mismo nombre que el perro de las tres y cuarto (visto de frente). Su propia cara en el espejo, sus propias manos, lo sorprendían cada vez. […] Había aprendido sin esfuerzo el inglés, el francés, el portugués, el latín. Sospecho, sin embargo, que no era muy capaz de pensar. Pensar es olvidar diferencias, es generalizar, abstraer. En el abarrotado mundo de Funes no había sino detalles, casi inmediatos.[[83]](#footnote-83)*

If a memory were produced by a single trace left by a single event, the subject would only remember particular events, like the chronicles of the Annals, being incapable of gathering them in order to make generalizations about them and form universal statements. Bernecker identified a solution to this problem: in his account, memory is capable of condensation, so the subject wouldn’t have the same problem as Funes because he is capable of generalizing. However, these generalizations are very limited and don’t enable us to explain the ability of humans have to permanently recombine the information they know and remember facts[[84]](#footnote-84) whose content cannot be precisely located in one or more past representations.

Returning to the topic of the new attempts to reformulate the causal theory, we can state that the idea of giving more flexibility to the content of the memory representation has led some philosophers of memory to give more flexibility to the trace as well, which has started to gain some independence of the event that produced it, getting prominence out of the past event in the causation of the memory as well as in the determination of its content. I will briefly analyze two of these recent reformulations of the causal theory: that proposed by Matthen, and that proposed by Michaelian.

*Matthen’s attempt: new content but same old trace*

Although the central point of his article[[85]](#footnote-85) is epistemological, it’s easy to see that Matthen conserves the basic schema inherited from the traditional causal theory of memory at the same time as he tries to incorporate the possibility of transforming and adding content. The importance given in his account to the memory trace is clear: what is preserved in memory is not the representational content or the belief or the experience itself, but the trace, which is always continuously occurrent and present, “whereas the beliefs [and images[[86]](#footnote-86)] in question are dispositional, and only intermittently occurrent”[[87]](#footnote-87). The content of a current memory depends not only on the corresponding memory trace but also on some additional information. Consider the examples cited by Matthen for both non-imagistic memory and episodic memory:

(A) If I hold the belief:

1. Sam is the executive chef of Lumière.

But someone told me that the chef of Lumière has changed, and I believe in what this person says, this new information does not erase the old information. My new belief could be:

2. Sam does not cook at Lumière, but he used to.

This belief could be the result of the following argument:

3. Sam was the executive chef of Lumière.

4. Now he isn’t.

5. Therefore, Sam used to be the executive chef of Lumière.

The content can’t be what is preserved in order to pass from the belief 1 to the belief 2 or 3 or 5 because clearly the meaning has changed. According to Matthen, “the continuing memory trace is (or is a part of) the categorical basis from my changing disposition to assent to propositions such as 1, 2, 3 or 5, and the proximate cause of the occurrent belief when they arise.”[[88]](#footnote-88)

(B) For episodic memories, Matthen points out that for some cases of chronological dating of an old event we have to associate it with other events that we can date directly. For example, “I learned a few years ago that an election was taking place in Iran. By now I don’t remember the date of the election, but I remember that I was aware of it happening. (…) To date the Iranian election, I need to remember not just that the election took place, but also my awareness that the election was taking place. (…) For instance, I might think: 'I remember reading about the election while thinking about my move to Toronto, so it must have been shortly before 2006, when I moved.'”[[89]](#footnote-89)

Leaving aside two interesting points mentioned by Matthen in relation to Fernandez, that is, the old debate initiated by Russell about the presence or absence of a feeling of pastness that accompanies an episodic memory and the possibility or impossibility of representing the past image from either the allocentric or the egocentric perspective (a possibility that is admitted by Matthen and so constitutes another argument in favor of the idea of additional information-processing in order to obtain the representation resulting from retrieval), we would like to point out that, even if Matthen is not directly concerned with the problem of the trace, it is not clear:

- first, what is the nature and role played by the so-called “additional information” which is not defined. In the first example, the “additional information” that exceeds the information provided by the trace seems to come from other trace left by other events (my friend telling me that the chef of Lumière has changed) plus a process of reasoning (which might well be unconscious). In the second case, however, the “additional information” does not become a part of the content of the memory but is only used as a cue (a kind of internal cue) in order to retrieve the desired event;

- second, whether the trace produced by an event can be modified or not. Given the examples discussed above and in the general thrust of the article, it is more probable that the trace remains unchanged over the time and that the additional information necessary to modify and enrich the content of an act of remembering is always external to the trace.

As we can see, although Matthen intends to assign a predominant role to the memory trace and not to the past event in the account of memory, he does not succed in going much beyond the traditional causal theory, due to a large extent to a static and immutable conception of memory trace.

*Michaelian’s account: A trace that changes but just a little*

On this score, Michaelian is more innovative, perhaps because he takes into account psychological studies. In his article “Generative Memory” he explores the implications of the psychology of constructive memory for the causal theory of memory. He rightly recognizes that construction doesn’t necessary imply content generation and he examines the constructive nature of memory at encoding and, at consolidation and reconsolidation, and its reconstructive nature at retrieval.

Concerning the encoding process, he adopts the distinction made by Alba and Hasher in four stages:

“Selection: Only certain incoming stimuli are selected at encoding.

Abstraction: The meaning of a message is abstracted from the syntactic and lexical features of the message.

Interpretation: Relevant prior knowledge is invoked.

Integration: A holistic representation is formed from the products of the selection, abstraction and interpretation processes.”[[90]](#footnote-90)

The reconstruction in retrieval is also taken from Alba and Hasher:

“Reconstruction: Whatever information was selected for representation and is still accessible used, together with general knowledge, (roughly) to generate a hypothesis about what might be happened.”[[91]](#footnote-91)

The interesting point that Michaelian notes is that at the moment of the reconsolidation of a trace, that is, the moment between a retrieval and a new storage, the memory trace is malleable and need not to leave unchanged. This seems to suggest that each memory representation of an event can introduce modifications into the corresponding memory trace. We could ask nonetheless if this representation would have to be a conscious act of remembering for the subject or if any kind of representation of the past event for any kind of use (not only for remembering) could also result in transformations of the memory trace.

Construction and reconstruction being essential to the nature of memory, content generation can occur at either of the stages mentioned: “content other than that provided by the initial representation might be incorporated into the memory trace before it is retrieved; and content other than that provided by the memory trace might be incorporated into the representation resulting from retrieval.”[[92]](#footnote-92)

This idea leads Michaelian to reformulate the traditional postulate of identity or similarity of content to the following one:

1. The content of the later representation does not go too far beyond that of the trace.

2. The content of the trace does not go too far beyond that of the initial representation.

The new formulation of the causal theory of memory that he proposes is the following:

“S remembers P iff

1. P;

2. S represents P;

3. there is a causal chain running back from S's current representation of P to an earlier representation of hers;

4. the causal chain goes continuously via a (distributed) memory trace with the content P (or something sufficiently close to P);

5. the causal chain goes continuously via a reliable memory system (responsible for the (re)construction of the trace and the current representation);

6. S's earlier representation had the content P (or something sufficiently close to the content of her memory trace); and

7. there is an appropriate relation between P and S's earlier representation.”[[93]](#footnote-93)

Nevertheless, even if he makes some interesting changes to the traditional theory of memory, which includes the consideration of the influence of prior knowledge on encoding and also on retrieval, the possibility of changes into the content of the memory trace from each retrieval/reconsolidation of it (which, moreover, is considered as distributed) and the requirement of a reliable memory system, the basic structure and problems that underlie his theory are the same, without counting the additional problems that arise from:

a. the use of the undefined formula “does not go too far”, which doesn’t really mean anything;

b. the idea of “content generation” which, even if its introduction is well-intentioned, in the sense that it’s a positive answer to the reductionists ideas of identity of content or similarity *à la* Bernecker, is so unspecific that it makes us think of creation *ex nihilo* of the content[[94]](#footnote-94). If this is not the case, which it probably is not, we can reasonably ask where the additional content comes from (for example, from other memory traces, from an inferential process?). Furthermore, we can also ask, given the idea that the content can be incorporated into the trace as well as into the representation emergent from the retrieval process, what is the nature of this content and wheter it is the same in both cases. As we can see, this notion is as vague as the “additional information” invoked by Matthen;

c. the notion of a trace whose content could be modified after a period or different periods of retrieval and posterior reconsolidation, which lead us to ask to what extent the content of the trace cannot “go too far” beyond that of the initial representation of an event in order for the trace to continue to be a trace of that event. Furthermore, we could also ask the old metaphysical problem known as “The ship of Theseus” about identity and change: in virtue of what does a trace of an event whose content is modified (maybe not only from different reconsolidations but also from other events or traces of those events) continue to be the trace of that same event?

d. Another question that we could ask is to what extent the actual knowledge of the system is really invoked to “generate a hypothesis about what might be happened” when the content of the memory event can’t really be different from the content of the past representation of that event.

As we could see, even if the idea of a trace that is malleable and can change is very innovative with regard to the idea of a static and unchangeable trace, it is, unfortunately, not exempt from some theoretical problems, especially because it continues to be irrevocably tied to the single past event that produced it.

It seems that Matthen’s and Michaelian’s problems are inverted: if Matthen allows the integration to the memory representation of a different content from that provided by the correspondeing trace but conceives of the latter as unchangeable[[95]](#footnote-95), Michaelian considers that the trace could be and generally is subject to some modifications (especially after retrieval) but maintains the existence of a strong correspondence between the content of the past representation, that of the trace and that of the memory representation, as well as a linear causal link between the three. So the real problem seems to be not only a static conception of the notion of a trace but principally the idea of a one-to-one correspondence, either between:

a- past events and memory representations;

b- past events and traces;

c- traces and memory representations.

**General problems of the causal theory of memory**

We can conclude that, even if these new attempts to reformulate the causal theory of memory are more flexible than the traditional account beginning with Martin and Deutscher and continued recently by Bernecker, all of them involve more or less the same assumptions, and thus give rise to the same problems:

**EPISTEMOLOGICAL**: Memory is conceived of as factive, at least at the moment of the first representation[[96]](#footnote-96);

**CONCEPTUAL**: We subdivided this item in the following aspects:

- FORM AND FUNCTION OF MEMORY: Memories assume a propositional form, or sometimes an imagistic form[[97]](#footnote-97), but never other kind of form, for example a discursive one; as I have already noted quoting Malcolm, this is a common postulate in the philosophy of memory, a product of an overintellectualization of memory. Memories seem to range over a huge spectrum of cases that go not only beyond the propositional format but also beyond the representational one. As Malcolm points out: “Suppose you called on a friend to take him for a drive and when the two of you left his house you saw him lock the front door. […] Your friend asks you, 'Did I lock the front door?' […] In one case you say, 'Yes'; in another you say, 'You locked it'; I another you lock the door yourself […] In one case you try to remember; but not in the rest. In one case you had a mental image of your friend bending over the knob; but not in the others. In some of the cases you had various feelings and thoughts; but none of these were present in all of the cases. Which of the various utterances, thoughts, feelings, or actions was the remembering or the memory event?”[[98]](#footnote-98) The point that I want to underline is that the representative function of memory does not seem to be the only function of memory, even if it was and is the only function considered in philosophy, in accord with a common and general trend to give an exacerbated importance to language, reference, truth and correspondence with reality. This reductionism is also reflected in another choice: the act of remembering, which corresponds to the conscious act, where the subject claims to remember something, has been the privileged object of philosophical studies, which have thus ignored the analysis of the vast spectrum of uses of memory that go beyond the explicit act of remembering. It’s unfortunate that we have forgotten for example what Bergson or Kierkegaard taught us, that is, that memory is not only used for contemplation of the past but also for present perception, present action and for future projection[[99]](#footnote-99). We could also add the neglect of the study of the relation between memory and reasoning, which could clarify the uncertain distinction between inferential and non-inferential memory.

- PASSIVE SUBJECT/STATIC MEMORY: We can find in these accounts two characteristics that are more or less related. On one hand, the subject who remembers is conceived of as a passive agent, and not as an agent with intentions, goals and different motivations, an agent who can also monitor and control (not necessarily in a conscious way) the reporting of his memories; in this sense, the integration of the notions of metamemory and of an active agent would be indispensable in a philosophical account of remembering. On the other hand, these causal accounts do not have a dynamical and holistic conception of memory which would inevitable imply that memory changes a little more than what is contained in the notion of reconsolidation (which on the other side only Michaelian takes into consideration).

- CONTEXT: the role played by context in these causal theories. As we’ve seen, the role played by context is generally neglected in these causal theories: only Bernecker integrates it in his account, but the role context plays there is minor: it only helps to determine the extent to which a current representation, in order to count as a genuine memory, can differ from the original one. All of the causal accounts discussed forget not only that the context of retrieval is also a significant factor in the determination of the content of our memory representations, but also that memory “operate on the internal wing of the vast, extended system of cultural and personal habits, hints, and patterns through which the inner representational regime has been sculpted and disciplined.”[[100]](#footnote-100)

**ONTOLOGICAL**: Under this heading, I include:

- FACTS AND EVENTS: The general assumption that in the world events and facts are intrinsically different from an ontological point of view (when they could be relative to the subject), with the plus that they are viewed as having clear and absolute temporal limits (which could also be relative to the subject and the context, or perhaps to the language and its narrative aspect). Besides, the causal accounts described above suppose that facts and events were represented in the past in a propositional or imagistic form, which suggest that they were consciously represented[[101]](#footnote-101). Bernecker assumes this explicitly (even if he denies, not very convincingly, that the representation is the object of the memory); because the other authors state that there was a past representation without giving either a definition of what is considered as a representation or an analysis that goes beyond the propositional or imagistic one, I conclude that they also implicitly accept this assumption.

- TRACE: The ontological status of the trace is not clear. Some causal theories don’t even specify it: we do not know if it is just a philosophical construct to explain the temporal gap, or a psychological entity or a neural one; others (e.g., Bernecker’s theory), conceive of the trace as having a double nature: mental and physiological, but this only generates other problems, as I have already pointed out. Besides, the trace is generally conceived of as an independent and static unity which, once it is fixed, does not change over time. Michaelian seems to be the only one who conceives the possibility of a mutable trace, but according to him, it can only change to a certain degree because it always has to be tied to the single past event that produced it.

- CAUSALITY: None of the accounts discussed clearly define the kind of causality concept that they invoke; nor do they say whether the concept is always used in the same sense: How is it possible that a past event cause a physiological memory trace? And is this causation similar to that which links the past event with the memory representation? Intuitively, it does not seem so. It is impressive that none of the defenders of the causal account of memory perceive the nature of causality as a problematic point, or at least as a topic to develop, preferring to rely on a naïve conception of causality, when even in ancient Greece philosophers distinguished between different kinds of causality.

- ISOMORPHISM, SIMILARITY AND CORRESPONDANCE ONE-TO-ONE: As Malcolm points out[[102]](#footnote-102), to postulate that two domains or things are isomorphic or similar demands defining the simple elements that compose each domain, in order to establish a key that permits comparing them. However, none of the causal accounts do this: what are the simple elements of the past experience? And those of the trace? Of the memory representation? The problem is that we can’t establish a kind of correspondence or correlation between a pas event, a trace and a memory representation. And this is because it’s difficult to conceive of an absolute sense of simplicity and complexity: as Wittgenstein points out “one and the same thing may be called simple or complex, depending on what comparison one is making –nothing is simple in an absolute sense, but that simplicity (and complexity) are relative to human choices and human ways of dividing up things”[[103]](#footnote-103); therefore, it’s impossible to devise a principle or a series of principles that could provide a reasonable criterion to establish a one-to-one correlation between a past event or representation and a trace, neither between a trace and a memory representation, or a past event and a memory representation, and even more difficult between their contents or structures (which have a different nature)[[104]](#footnote-104). Setting aside the impossibility of explaining how the trace conserves the information, Bernecker tries to determine some principles (which, by the way, are quite reductive) to compare the past and the memory representation; however, he fails, not only because he reduces this analysis to propositions, but mainly (and like the others authors) because of presupposing that these propositions are the simple elements that have to be compared in order to determine if it is or is not a genuine memory representation. On this point, Malcolm is enlightening: “At 8 A.M. you 'walked out the door', you 'crossed the threshold', you 'stepped like this', you 'left the house', you 'began walking toward your car', you 'started for your office', you 'began another working day' and so on. What you did at 8 A.M. could be related to many different contexts, thus yielding many different descriptions of what you did. It would be absurd to think that there are only ten or fifteen true descriptions of what you did at 8 A.M. The number is indefinite because the number of contexts and concepts in terms of which your action might be described is indefinite. But then, what you remember doing at 8 A.M. is, in exactly the same way, indefinite.”[[105]](#footnote-105) We can easily reformulate this example to take into consideration “facts” learned, like the second World War. I can remember that:

(a)“The Second World War started in 1939 and ended in 1945”;

(b) “The Second World War started with the German invasion of Poland”;

(c) “During the Second World War started in 1939 Nagasaki was the target of the world's second atomic bomb attack”;

(d)“The Second World War, the biggest massacre of the twentieth century, started on 1 September 1939 with the German invasion of Poland, and ended with total victory of the Allies over Germany and Japan in 1945”;

and so on; we can even imagine a discourse and many different discourses about my Second World War memories. Is it reasonable to think that each of these acts of remembering is caused by a different fact that I learned via a single and different memory trace? Maybe Michaelian and Matthen would accept that (a) and (b) are caused by the same memory trace, but what about (c)? In their accounts, it would be caused by another memory trace different from that one corresponding to (a), even if both have in common a small piece of information: the date of the beginning of the war. The same happens to (d), which seems to incorporate a lot of information, some of it present in the other memory representations, some of it not. Would (d) be the result of a single memory trace? And what would happen if I learned the information contained in this proposition at different times? There would clearly not be a past representation that could be understood as the cause of my current representation. So would (d) be the result of several memory traces, each one caused by a single fact learned? This would be impossible not only because the explanation would involve an unnecessary multiplication of traces (which would also be needed if we were to suppose that each proposition is caused by a single memory trace), but principally because we can’t determine which would be a simple unit of information that would correspond to a single trace caused by a past representation of a single fact. The same problem would arise if we express what we remember not only in a single proposition but in a discourse. Even more, it would seem more likely that “semantic memories” are organized by topic, independently of the specific past events that causes them: for example, if I remember (a), would the cause of this memory be the first time a teacher mentioned this information? Or, when I later read about it in a book? Or when I studied for the exam and repeated these dates time and again? Which specific event of the past would be the cause of me having this semantic memory of the Second World War?

In conclusion, it seems that the correlation between the past experience, the trace and the memory representation of it, the correlation characteristic of the causal theory of memory, don’t provide a good account of how memory works in the real world. Maybe it is useful for explaining some cases of memory, for example when we learn items or propositions by heart and then have to remember them; but these cases are more common in psychological laboratory experiments that in real life. This doesn’t mean that the causal theory as it has been developed doesn’t have an explanatory power; the problem it is that its field of action is very limited. We can compare this case to what happens in the physical field: as we know, the quantum mechanics and the general relativity are both exact in their field of application; the first one explains the smallest objects of the universe and the second one the biggest ones, but each is incapable of explaining the objects explained by the other. String theory, even if it is not fully developed, presents itself as capable of explaining both objects: the infinitely big and the infinitely small. Maybe it’s the time to search for kind of string theory for memory.

**SECOND CHAPTER: PSYCHOLOGICAL MODELS OF MEMORY**

In the first part of this essay, I discussed different versions of the causal theory of memory. In this second part, I am going to present some cognitive psychological models of memory. My purpose is to analyze how the phenomenon of memory is understood in these different models, in order to answer two questions:

(a) whether we can also find the assumptions typically made in the causal theory of memory or not;

(b) whether there is a some model that promises a totally different conception of memory, which would provide another argument (or arguments) against the causal theory and hint at a different formulation of a philosophical theory of memory.

I will develop three different kinds of psychological models:

- the spatial analogy, which is really a naïve model;

- two symbolic models of memory;

- two connectionist models of memory, one local and another distributed.

**The spatial analogy**

The spatial analogy is the most common and ancient conception of memory, developed mainly by the British empiricist philosophers through the conception of an associative-atomistic mind as a store of elementary ideas and associations between them. This metaphor of memory has long dominated most memory research and theory. The basic assumptions are the following:

a. memories are considered to be discrete, static objects stored in particular locations in mental space, each one representing a complete experience;

b. the salient characteristic of memory is its countability; thus its performance is measured in terms of the number of recovered items;

c. this last feature entails a conception of forgetting as information loss;

d. the use of metaphors is frequent: the most common is the metaphor of memory as a storehouse (in some cases, organized in many departments). However, this traditional metaphor generated a variety of related species, like a library with books, a dictionary with words, the computer analogy, etc[[106]](#footnote-106);

e. there is a clear distinction between the mental content and the mental processes, that is, between the information stored and the operations upon it; because memories exist independently of the operations of encoding, retaining and recalling them, it’s quite natural to conceive of the process needed to retrieve the information as an “executive” or “homunculus” that must look through the various memory stores in order to find the information required;

However, these spatial accounts (or, better, since they do not constitute real accounts, spatial analogies), do not differentiate between the memory event and the memory trace: what is stored in memory is exactly the same as what is directly retrieved by the executive processes in the act of remembering and presented as the memory of *x*.

This simple metaphor shaped and unfortunately continues to shape most of the phenomena studied by laboratory memory research, including the effects of list length, retention interval, spacing, serial order, etc. It continues to be used even in connection with memory phenomena for which it was not originally intended and which seems to call for a different type of explanation, for example, the case of autobiographical memory, which clearly is not a simple set of remembered concrete experiences; even if there are “horizontal” links between successive events, autobiographical memory seems to have an hierarchical structure “rich enough in overlapping and nested relations”, as Neisser explains[[107]](#footnote-107).

**Symbolic models**

Following the classification proposed by Roediger[[108]](#footnote-108), we have to distinguish between the spatial analogies mentioned above, which are simple metaphors, and more elaborated memory theories that nonetheless have a spatial overtone, which we’re going to develop next. In general, the main differences are:

(a) In contrast to point “a” mentioned above, in these theories the memory trace is not a discrete object stored in the mind that represent an experience wholly, because each experience is decomposed by sensory processes into distinguishing features, which are stored as different aspects of the memory trace of the event, in a language-like form (in symbolic models) or in a network of nodes (in local connectionism): “memory can be represented, much as in the subway-map model, as a great network of nodes that are linked by associative paths. Experiences are encoded by making appropriate nodes and their connecting associative paths. Retrieval consists of searches through the associative network for the desired information”[[109]](#footnote-109). But even if the trace can’t be compared with a particular object, and in this way it could be considered distributed between multiple units, these models are classified as localist models, because they continue to establish a one-to-one correspondence between the minimal unit and a concept.[[110]](#footnote-110)

(b) These accounts are not only concerned about quantity, that is, with “how much” is remembered, but especially about the quality of memory, with “what” is remembered. This is emphasized because the feature that is considered to be essential in memory is the accurate representation of the past, that is, the correspondence between the memory representation and the past event. Because memory is mostly conceived as propositional representations, a truth value is in general attributed to memory reports; they “can be judged as right or wrong, or as being more or less 'true' to aspects of the actual event.”[[111]](#footnote-111)

(c) This last feature entails a conception of forgetting not as a loss of items but principally as a loss of correspondence between the memory report and the past event, i.e., as a deviation from veridicality.

(d) As we’ve already mentioned, memory traces are generally represented as a network of units interconnected by associated paths, like a subway-map model.

(e) In a sense slightly different than the spatial metaphors, mental content and mental processes continue to be different: on one hand, there are structures of association between the units, and on the other hand there is the activation of these associated paths. The memory trace and the memory event are also the same, as in the spatial metaphors: at recall, traces are simply reactivated.

*Quillian’s model*

The characteristics mentioned above are typical of symbolic models, for example, that proposed by Quillian. Quillian’s theory of semantic memory is of great importance because it was one of the first examples of a network model of semantic memory and because Quillian was also responsible for the current resurgence of work on “spreading activation” (an idea which is also essential to connectionism)[[112]](#footnote-112). It’s important to note that Quillian’s theory was first developed as a theory for digital computer, which tried to show that machine translation programs only could solve certain problems, in particular the resolution of polysemy, “by storing the meaning of natural language words in a medium and a format providing properties similar to those of human 'understanding'”[[113]](#footnote-113) Only later did Quillian propose his theory of computer memory as a reasonable simulation model of some aspects of human memory[[114]](#footnote-114). The question Quillian tries to answer is: how is semantic information organized in human memory? His theory focuses especially on the analysis of word meaning, which clearly differs from dictionary definitions. The principle that guides memory, according to Quillian, is avoiding redundancy, in such a way that, for example, there is no need to restate the information constituting the concept of “machine” in the information the subject has as the meaning of “typewriter”. In this model, memory is conceived as a mass of nodes interconnected by different kinds of associative links. Each node corresponds to what Quillian calls a “property”, which can be a mental picture or a language concept, even if Quillian only analyses the cases of language concepts[[115]](#footnote-115). There are two kinds of nodes: type nodes, whose associative links may lead directly into a configuration of other nodes that represents the meaning of its name word; and token nodes, which refers indirectly to a word concept by having one special kind of associative link that points to that concept’s type node. For any word meaning there can be one and only one type node in memory, but there will in general be many token nodes scattered throughout the memory. So the general structure of the memory consists on what Quillian calls “planes”, which, in this semantic memory model, correspond to the full concept that the subject has of a word, which are made up by a patriarchal type node and all the type nodes that are linked to it. From each of the token nodes linked to a type node, there are links to other type nodes, and from each of these in turn to still others. On the other hand, there are also different kinds of links. Quillian distinguishes five cases:

- superordinate and subordinate links, whose nature is hierarchical. The first link is in general known as the ISA link;

- modifier links, which indicates a modification of the type node by any word or phrase used adjectively or adverbially;

- a disjunctive link, required specifically by the multiple meaning of the type nodes;

- a conjunctive link, required, for example, between some token nodes within a plane;

- an open-ended category of links, by means of which all the remaining kinds of relationships are encoded, which is necessary because it’s not possible to specify in advance the enormous and various kinds of relationships that can be formulated in natural language.

The following schema represents a hypothetical memory structure of three-level hierarchy from the category “animal”. Information true of animals in general (such as the fact that they eat, can move around, breathe, have skin, etc) need not be stored with the memory node for each separate kind of animal. So, for example, since all the living things can breathe, the proposition is stored with “living thing” at the top of the branch; but since only birds can fly, the proposition is stored with “bird”. One property of the model is that it permits adding new information and also making inferences simply from the addition of a proposition: for example, from new information like “an ostrich is a bird”, we can infer that it can breathe, can move, has feathers, etc.

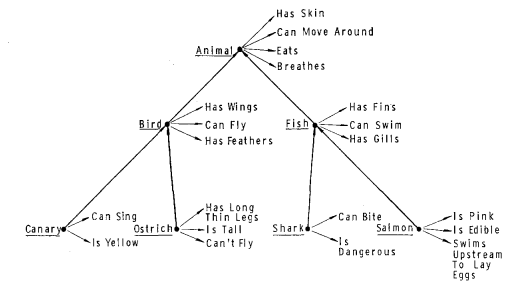


Fig. 1: Illustration of the hypothetical memory structure of a three-level hierarchy[[116]](#footnote-116).

However, this model deals neither with acquisition of semantic knowledge nor with recall memory, but only with recognition of word meaning. So it can explain from the notion of “spreading activation” through the nodes (first to all the nodes linked to the first node, then to all the nodes linked to each of these nodes, and so on, until find an intersection between two nodes), that for example when hearing or reading a phrase such as "the fall leaves" a path found between the concept "to fall" and the concept "tree leaf" would be rejected as a wrong interpretation, and other paths would be generated until found the one which would be accepted as correct. But it doesn’t explain how new information is incorporated.

This model nonetheless has been criticized because it does not explain data from a range of experimental tasks and manipulations[[117]](#footnote-117), for example, the fact that people do not take longer to access to general properties than specific ones; sometimes, general category properties can even be verified more easily than more specific category information[[118]](#footnote-118). Furthermore, this exclusive concentration on semantic aspects of language assumes that the most important aspect of a word is its dictionary definition, and that the only aspects stored in LTM are the semantic ones; other features of language are not, being only briefly stored in STM. Nevertheless, empirical research proves that sometimes superficial and irrelevant features of language (accentuation, pronunciation, typography, color, voice, etc) and other contextual details (like the pragmatic use of a word) are actually remembered with the words or serve as cues to remember the words. Clearly this cannot be explained by this model. And we could question too if it is an actual economic model of the semantic aspect of the words.

On the other hand, this model only tries to give an account of a particular part of what is considered “semantic memory”: word’s meaning. But what about the other information that usually enters into the “semantic memory” category, like facts? Can they be integrated into this node net of meanings? If so, how is it possible? And what of the “episodic information” that sometimes seems to be encoded with a fact of the world, for example, some aspects of the context of learning? As we can see, this model is a very reductionist one. What is interesting to note is that in spite of its reductionism, the idea of organizing knowledge of word meaning around a type node that represents a concept is clearly inconsistent with the idea in the causal theory of memory of an univocal correspondence between a fact learned, the trace of it and its memory. This model suggests that traces about semantic information are, on one hand, organized thematically, and on the other hand, susceptible to modification, and so not really causally tied to previous representations. For example, is it plausible that one single past representation was the cause of the memory trace that allows a person to know (that is, to remember) that birds can fly? Doesn’t it seem more plausible that a person learns this information in childhood from various experiences; i.e., seeing birds flying, adults telling him that these animals are called “bids”, seeing other animals that do not fly and progressively learning the meaning of “fly”, etc.? What is impressive is that this idea about the acquisition of semantic information is compatible with Quillian’s model, which was elaborated in the sixties, but not with the causal theories of memory that continue to be proposed today. It would be interesting to go more deeply into the experimental psychological studies of child’s acquisition of basic knowledge of the world; this would really help to clarify if a biunivocal correspondence between a representation of a past information and the memory of it is likely or not.

Other symbolic architectures were developed following Quillian’s model. For example, the Search of Associative Memory (SAM) model developed by Raaijmakers and Shiffrin[[119]](#footnote-119). The SAM model can deal with many more phenomena than Quillian’s model, like free recall, paired-associate recall and recognition. It has the advantage of considering memory retrieval as cue-dependent and thus, that temporal-contextual information is of fundamental importance; that is, in this model, what is stored in long-term memory is not only word information, i.e., information that enables the subject to produce the name of the encoded word, but also word-context information, which refers to temporal and situational factors like “environmental details, physical sensations, emotional feelings, and all thought processes not directly relevant to name production”[[120]](#footnote-120). However, a clear disadvantage is that this model only analyzes data in terms of the number of recalled words or the probability of recalling them.

*The ACT theory*

On the other hand, there are more elaborated models that try to explain not the organization of word’s meaning but the organization and also the acquisition to some extent of declarative information in general, that is, of semantic and episodic information understood in a broad sense, like the ACT theory. The ACT theory (Adaptative Control of Thought) was mainly developed by Anderson in different versions (1976: ACTE, 1977-1980: ACTF, 1983: ACT\*, 1993: ACT-R). Next, I will focus my attention on the ACT\* version, for two reasons: first, because the distributed connectionist model that I will present later was created in that time period and was meant to differ in certain respects from this version of ACT, and second, because the only difference is that in the ACT-R model the principles for activation and for conflict resolution are explicitly guided by the rational analysis of cognition, although this difference is not decisive for our purposes. We can summarize its basic postulates as follow:

(a) The unity of the principles that explain higher-level cognitive functions like language, reasoning, problem solving, and also memory;

(b) The assumption of a set of condition-action pairs called productions that underlie human cognition: “The condition specifies some data patterns, and if elements matching these patterns are in working memory, then the production can apply. The action specifies what to do in that state. The basic action is to add new data elements to working memory”.[[121]](#footnote-121) For example,

IF person 1 is the father of person 2

and person 2 is the father of person 3

THEN person 1 is the grandfather of person 3

is a typical production rule that would apply if, for example, “if Fred is the father of Bill” and “Bill is the father of Tom” were active in working memory, which would allow the inference of “Fred is the grandfather of Tom” and its later deposit in working memory.

(c) Three memories: working, declarative, and production memory:

- working memory: this contains the declarative information that the system can currently access, which consists of the active information retrieved from long-term declarative memory as well as temporary knowledge structures deposited by encoding processes and the action of productions;

- declarative memory: because the declarative form is the first form in which knowledge comes (which must not be confused with the propositional form), declarative memory contains all the information stored.

- production memory: it contains the productions mentioned before. The declarative form in which knowledge comes is in some cases interpreted by general procedures by performing a task called *proceduralization*, which gradually replaces the interpretative application with productions that perform the behavior directly. For example, rather than verbally rehearsing the side-angle-side rule in geometry and figuring out how it applies to a problem, a subject eventually can have gradually acquired a production that perform the behavior directly. So, even if this is not clear, it would seem that this memory includes more than the classical procedural memory; it is also concerned with all kind of “cognitive skills”, including heuristics, cognitive shortcuts, schemas, etc.

(d) Four processes (ignoring encoding and performance processes both of which are not central to the ACT theory):

- a storage process: this can create permanent records in declarative memory of the contents of working memory and can increase the strength of existing records in declarative memory;

- a retrieval process, which retrieves information from declarative memory;

- a match process: data in working memory are put into correspondence with the conditions of productions.

- an execution process: it deposits the actions of matched productions into working memory.

The combination of these last two processes is referred to as production application, whose arrow in the schema cycles back into the production memory box, reflecting the fact that new productions are learned from studying the history of application of existing productions. The introduction of this notion demonstrates clearly a more dynamic conception of memory as well as some idea of a learning memory system.

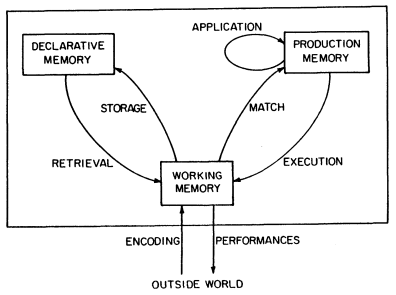


Fig. 2: A general framework for the ACT production system, identifying the major structural components and their interlinking processes[[122]](#footnote-122).

Another interesting point is that according to Anderson, even if ACT\* is a rule-based system, it is not incompatible with a neural system that implements the rules: both of these are necessary to an accurate psychological theory because they represent different levels of description. For him, those who defend only one system are just confusing different levels of explanation[[123]](#footnote-123).

In the ACT\* theory, memory is understood in terms of the network structures that encode knowledge. The units or chunks in which knowledge comes and on which memory processes operate are called “cognitive units”, which consists of a unit node plus a set of elements encoded in a particular relationship to it. For example, in the case of propositions, the unit node is the proposition itself, which generally consists of a subject, a verb and an object, whereas the other complements are the relation and its arguments. The number of elements is not infinite; Anderson limits it to five. In order to generate the complex network that is knowledge, the cognitive units are encoded in a hierarchical structure: smaller units serve as elements of other larger units; elements occur in several units, creating subhierarchies that Anderson calls “tangled hierarchies”, for example, when a subproposition or an image, like an image of a person, can appear in multiple propositions, encoding for example different facts about that person. I will give some examples later. I will now simply not a brief remark made by Anderson: subjects do not necessary adopt an entirely consistent hierarchical encoding scheme.

The cognitive units are also called “representational types”, because they differ in their representational assumptions. In the first model, ACTE, Anderson held that a cognitive unit could only be a proposition because knowledge was conceived exclusively as propositional; but in the ACT\* model he introduces three types of cognitive units: temporal string, which encodes the order of a set of items (for example, words, or numbers); spatial image, which encodes spatial configuration; and abstract proposition, which encodes meaning. These three types correspond to the three possible representational types of knowledge, which are defined not in terms of the notation that expresses them (which can be used in a similar way for the three types), but in terms of the processes that operate on them: encoding, storage, retrieval, match and execution processes. Storage and retrieval processes, that is, declarative memory, treat the different representational types in the same way: either all the elements of the unit are encoded and retrieved or none of them is. However, the other three processes which correspond to working memory (encoding, matching and execution) treat them differently. An example: in the case of strings, the encoding process preserves only the ordinal structure of the events, whereas in the case of spatial images, it preserves the configural information of elements in the spatial array about relative position and orientation, but not an absolute size or distance. I will not develop all the differences between the three representational types; I will focus my attention on the encoding and execution of the propositional representation.

Propositions are the most abstract of the three codes. The encoding process is guided by a principle of economy: the only thing that is encoded is that two arguments (subject and object) are in an *x* abstract relation (which corresponds to the verb). Surprisingly, Anderson specifies that the only kind of representational type that is not a reproduction but an abstraction is the propositional one: “unlike the encoding processes for temporal strings or spatial images, the structure of an abstract proposition is not a direct reflection of environmental structure. Rather it reflects an abstraction of an event, and the encoding process itself is something that must be learned”.[[124]](#footnote-124) People learn which aspects of a higher-order property of an event are significant, and in order to represent the significant relationship directly they develop a code, which allows them to avoid representing all the pieces of information or the exact words of the sentence. For example, instead of encoding all the following propositions: “A raised his hand over his head”, “A's hand held a round object”, “A's hand thrust forward”, “A's hand released the ball”, “the ball moved forward”, etc., people abstract the events into inferential proposition like “A has thrown a ball”, which is subsequently stored as the abstract representation of the those events. Another example of the principle of economy: the knowledge organization by theme nodes, to which thematically related predicates are associated. In order to represent a subset of facts about, for example, a person, subnodes are created to connect the traces of that subset with the cognitive unit corresponding to that person, as in representation (3b). As Anderson explains, “the subnode is basically a *token* of the theme node, which is the *type*.”[[125]](#footnote-125) So knowledge is organized in different abstract theme types, and facts and events are represented as tokens of them. The execution process corresponds to the creation of propositional structures from the combination of different elements, primitives or not; and also to the inference procedures that fill missing slots.

I will nonetheless present a schema of each representational type:

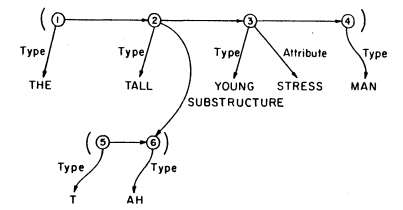


Fig. 3: A representation of a string network[[126]](#footnote-126)

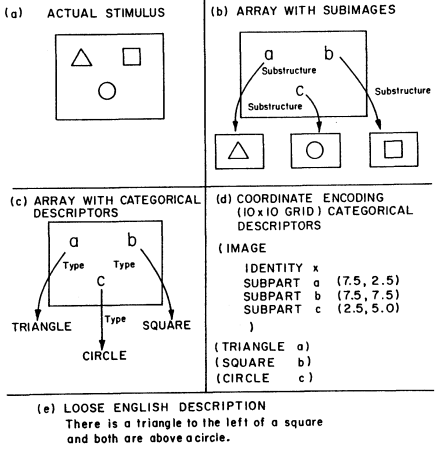


Fig. 4: Alternative notations of spatial images[[127]](#footnote-127)

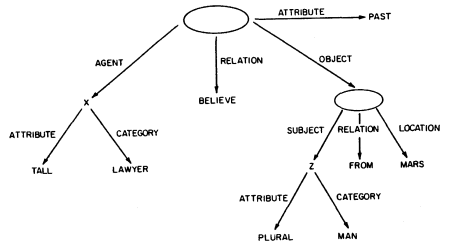


Fig. 5: A representation of a proposition encoding[[128]](#footnote-128)



Fig. 6: A network representation of two types-facts about a subject called Marty[[129]](#footnote-129)

As we can see from these figures, there isn’t necessary a single notation corresponding to each representational type; for example, in the representation of the spatial image (2), the visual detail can be encoded in subimages as in (b) or through categorical descriptors, like words in (c) and coordinates in (d). This shows that representations of a hierarchy don’t necessary have to be of the same type; they can be mixed, that is, multirepresentational, and they often are. For example, “to represent the sequence of events at a ball game, one might want a linear ordering of a sequence of propositions describing the significant events. Strings and images would be mixed to represent a spatial array of nonsense syllables or a sequence of distinct images”[[130]](#footnote-130). Or, as figure (4) shows, “a string encodes the main element of the restaurant sequence (enter, order, eat, and exit), another string unpacks the sequence involved in entering, an image unpacks the structure of a table, and so on”[[131]](#footnote-131).

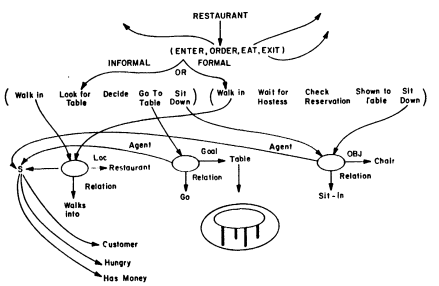


Fig. 7: A tangled hierarchy of multiple representational types[[132]](#footnote-132)

I will now present, in a schematic form, the basic processes of memory as they are conceived by Anderson in the ACT\* theory. I will nonetheless omit its formalization due to its complexity and lack of space.

- Encoding: When a cognitive unit is created in order to record some external event or the result of some internal computation in an active but transient state, it is placed first in working memory, and then with a certain probability is turned into a permanent log-term memory trace (concept reserved for permanent units), which will have one unit strength. If a cognitive unit is created in working memory and a copy already exists in long-term memory, its long-term strength will increase by one unit and a new trace won’t be created. This probability is an assumption of the ACT theory, and is considered independent of intention, motivation and duration of residence in working memory.

- Retention: Each node in declarative memory has an associated strength, which is basically a function of the frequency of use of that cognitive unit (which is clearly connected to the familiarity of the concept) and of time. The first successful trial establishes the trace with a strength of one unit; each subsequent trial increases the strength by one unit. The element nodes of the cognitive units also have their own strength: they increment their strength every time the corresponding unit node acquires an increment in strength. However, because an element node can participate in more than one unit node, it can have more strength than a unit it participates in. Thus all nodes in the declarative network have associated strengths, and all accumulate strength with practice. The trace once formed is never lost; what happens is that the strength of a trace can decay. The total strength is the sum of the strengths remaining from the individual strengthenings. The strength of a trace is important because it determines its probability and speed of retrieval. Furthermore, it affects its level of activation in two ways: first, it’s an indispensable factor to determinate the amount of activation spread to a node from associated nodes; and second, the strengthening of a trace produces an increment in the activation capacity of its elements.

- Retrieval: The probability of retrieving a trace and the time taken to retrieve it are functions of the trace’s level of activation, which at the same time depends on its strength. Activation, which is viewed as a kind of “energy” that runs the cognitive system, controls the rate at which information is processed by the pattern matcher for production conditions. Because working memory and long-term memory overlap in terms of their contents, the units and elements of the working memory (either an encoding of a perception of objects in environment or of the result of an internal computation recently processed) are the source of activation of the elements and units of the long-term memory, which are at the same time associated with other units: this is why activation spreads from the working memory to the long-term memory network. In the following example, two units in working memory, “The doctor hates the lawyer” and “The lawyer is in the bank”, encode the sentence “The doctor hates the lawyer who is in the bank”. The nodes corresponding to the main concepts are the sources of activation from which activation spreads throughout the long-term memory network, activating other units like “The doctor hates the priest”, “The lawyer hates the judge”, “The lawyer was in the court on Friday”, “The sailor is in the bank”, “The bank was robbed on Friday” “The sailor was robbed on Friday”, etc.

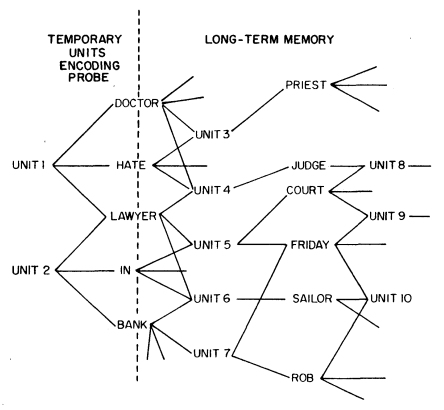


Fig. 8: A hypothetical network structure[[133]](#footnote-133)

Retrieval (or recognition, depending on the case) occurs when activation from various nodes and/or elements has intersected (which is monitored by the memory system) and its level has reached some threshold value. The levels of activation of the nodes in the network reflect their degree of association to the source nodes: ACT\* predicts that if two structures might be matched to the same pattern, the subject will prefer the more active one. Therefore, activation measures how closely associated a piece of information is to information currently used. For example, for the cases about specific information stored in theme nodes (as in representation 3b), if there is not a related foil subject, the subject simply retrieves the subnode, following its higher level of activation in comparison with the others; if there are related foils, however, the subject has to retrieve the target trace which corresponds to the specific fact, which is done in a two-stage process: first, the subnode is identified and then activation spreading from the subnode enables the target fact to be identified (this process is referred to as “refocusing”).

I want to add two other points which, even if they are essential neither to the ACT theory nor to my main interests, can help us to understand how traces are conceived in this theory.

The first point refers to what Anderson calls “elaborative processing”, that is, when the subject, in order deliberately to encode some information, creates an elaboration or choses an old one, which subsequently allows him to improve recall. Anderson describes three types of elaboration:

1. The multiplication of paths between the items to-be-remembered; for example, if the pair is dog-chair, the subject can generate the following elaboration:

The dog loved his masters. He also loved to sit on the chairs. His masters had a beautiful black velvet chair. One day he climbed on it. He left his white hairs all over the chair. His masters were upset by this. They scolded him.

which would generate the following network representation of the elaborative structure to connect the pair:

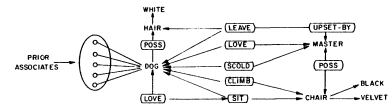


Fig. 9: Network representation of the elaborative structure generated

to connect the pair dog-chair[[134]](#footnote-134)

In order to discriminate between target traces and elaborations, Anderson assumes that part of the trace is a tag indicating whether it is an encoding of a study event or part of a subject elaboration.

2. The generation of additional concepts from which to spread activation to the target. For example, if the subject has to remember a sentence like “The minister hit the landlord”, to begin with, he might retrieve a schemata like “people hit people with hand-held objects”, that would lead to the elaboration that the minister hit the landlord with an object and that he held the object. This elaboration process can continue by retrieving another schema; for example, if he saw and recalls a movie where a minister carried a wooden cross when confronting the devil, this would lead to the elaboration that the hitting instrument was a wooden cross and that the landlord was the devil. The impact of such elaborative activity could lead to include some of these intersecting concepts, like for example *cross*, in the trace of the sentence, like it is represented in the following network representation:

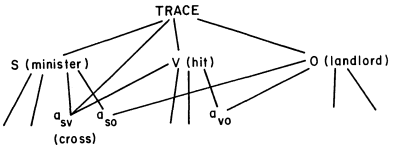


Fig. 10: Network representation of a subject-verb-object and the overlapping network of associations among the concepts[[135]](#footnote-135)

3. The use of schemata, that is, which can be a set of events already stored in long-term memory which can serve as analogs for elaborating the current event, or general events characterizations. For example, suppose the subject has to remember the following sentence “The janitor chased the cat” and has in memory the schemata derived from actual events in the subject’s life:

The janitor found the rat in the basement.

The janitor chased the rat.

The rat fled out the window.

Combined together the subject can construct the following elaboration by a simple substitution of terms:

The janitor found the cat in the basement.

The janitor chased the cat.

The cat fled out the window.

This elaboration allows him to regenerate the sentence he studied, even if he doesn’t remember it in the first instance; if the schema is retrieve, it can be used to generate the original study sentence that is being searched for.

The introduction of the notion of elaborations has the advantage of explaining on one hand, semantic intrusions and false alarms, due to the interference of the elaborations into the trace we want to retrieve; and on the other hand, the generation of a past representation richer than the original material, producing in this way an embellishment of it. For example, if the subject studies:

1. Fred christened the Tia Maria.

2. Fred broke the bottle.

3. The Tia Maria went to Vancouver'

4. Fred did not delay the trip.

and has in memory the following schema:

1. Person christens ship.

2. Person breaks bottle against ship.

3. Ship leaves port.

4. Ship goes to location as a trip.

At retrieval, he’ll probably produce the following embellished story:

l. Fred christened the Tia Maria.

2. Fred broke the bottle against the Tia Maria.

3. The Tia Maria leaves some port.

4. The Tia Maria goes to Vancouver as a trip.

5. Fred did not delay the trip.

The second point I want to mention refers to the incorporation of the context, in two senses: in some cases the context is encoded as an element of the cognitive unit, for example, for certain strings, as word recognition, where the subject forms traces linking the words to the various contextual elements. In the following representation, each line corresponds to a trace; the contextual element is represented as a single context node.

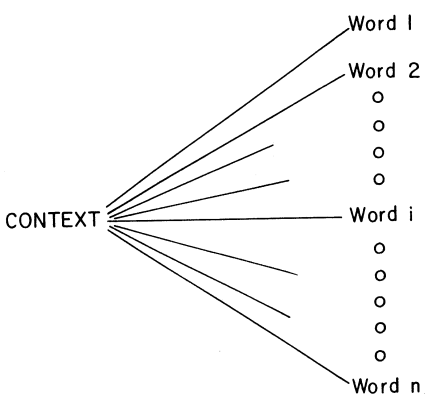


Fig. 11: Network representation of the word-context association for a single list[[136]](#footnote-136)

In other cases, the context is what determines the sense of an item which has multiple senses. If, for example, for the cases of memory for items, the word “black”, which is generally associated with “white” as referring to a prototypical color or a to a race of people, is also associated with “train” in an specific context of study, referring to soot or the color of a train, it’s the context of retrieval which is going to determine the sense chosen through the spreading activation in a double process: the wave of activation determines first the sense of the word, which later spreads the activation from the chosen sense in order to retrieve the trace:

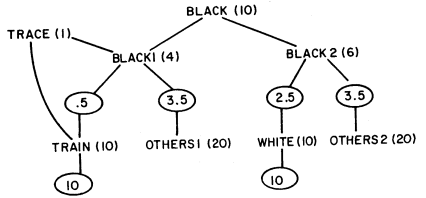


Fig. 12: Memory representation when the subject has studied the pair “train-black” (the numbers are the strength associated with the spreading activation)[[137]](#footnote-137)

The ACT\* theory accounts for a broad range of phenomena from the memory literature, which according to Anderson constitutes an evidence in favor of this mechanism, for example, the fan effect of interference for associated information, and the primacy and recency effects for list memory and serial recall. For the author, it’s even possible to explain by the same principles the encoding, retention and retrieval of more complex and richer structures like stories and discourse, though he doesn’t develop this point.

I will not discuss the viability of this model here, because this would also imply discussing the debate between symbolic and connectionist models, which far exceeds the scope of this work. I only want to note the following points:

- Even if most of Anderson’s analysis focuses on the propositional representation of declarative knowledge, it has the advantage with respect to other symbolic models (like that one of Quillian), of considering the existence of other types of representations that are not propositional, like spatial images and temporal strings, based on a processing criterion and not on a notational one. Moreover, the possibility of multiple types of representations within the same cognitive unit, although not very developed, seems very innovative not only relative to other symbolic models but also relative to the philosophical theories of memory that we presented in the first part of this work, where semantic memory is basically reduced to propositional representations, and if the imagistic form is considered in some cases, but in general for episodic memory (like what is generally called semantic memory couldn’t acquire an imagistic form), it is always thought as giving rise to a different kind of trace whose all nature is imagistic. As we can see, the simple idea of a multirepresentational trace seems to be absent of the philosophical theories of memory.

- Another classical dichotomy that seems to be questioned at least to some extent is the distinction between semantic and episodic memory, which in general is taken as a totally established distinction not susceptible to revision. In ACT theory, particular episodes, even if they constitute separate traces, are nevertheless intrinsically associated to semantic structures like some schemas[[138]](#footnote-138), which can even alter the content of the trace, producing what would be considered from an epistemological point of view a false memory, but from another point of view in some cases an intrusion or maybe a simple embellishment, that seems to escape from the truth value criteria. Even if the examples given by Anderson correspond more to laboratory research on memory and not to everyday memory research, and even if the ACT theory explains neither the acquisition of the general and abstract schemas (which constitutes a great gap in the theory -it would be interesting to know for example if episodic memories play a role in the generation of these schemas) nor the possibility or impossibility of their modification, the ACT theory at least denies the existence of a categorical boundary between episodic and semantic knowledge at the encoding, storage and retrieval moments, an idea very common on the philosophical memory field.

- Furthermore, even if it’s more insinuated than developed, context is considered not only as one factor that determines the direction of the spreading activation at the moment of retrieval, but also as information that is encoded in some cases within the same memory trace. Once more, causal theories of memory do not even acknowledge the latter possibility.

- Finally, the nature of the trace proposed in this account is also interesting, because this symbolic model incorporates the idea of spreading activation: traces are not exclusively the cognitive units but the connections between them, which are determined by their strength. Therefore, we could say that the nature of the trace will be a kind of mixture between these units that would be like a stable entities stored in memory (as in the spatial analogies) and the connections between them that are based on their strength and would have a kind of energetic nature.

**Connectionist models**

Connectionism has also presented some memory models. Even if the symbolic models we developed propose some interesting features of memory often neglected by philosophical accounts, they are too tied to the analysis of linguistic facts and relations. Connectionism tries to explain changes of state through more dynamical models, at least in distributed accounts, although it has had more impact on the modeling of what we can call “lower” cognitive processes than in “higher” ones, like recognition of words, the past of verbs, etc. We have to distinguish two kinds of connectionist models according to the nature of the minimal unit: when it represents a concept or a word or an attribute in a simple one-to-one correspondence between them, the model is considered *localist* and, as we are going to see, is very similar to symbolic models; when the unit does not stand for a particular concept or feature because it is a pattern of activity over the set of units that in certain way represents them, the model is considered *distributed*.

*McClelland’s model: local connectionism*

The classic memory model of local connectionism is that proposed by McClelland (1981)[[139]](#footnote-139): “The basic idea is that representations of previously-experienced exemplars stored in memory are activated via a spreading activation mechanism. Activated exemplars themselves activate representations of their properties. Mutually exclusive property values compete so that properties which are supported by a large subset of the active instances of the category are reinforced and become strongly active while those which are not are suppressed”[[140]](#footnote-140).

The example that McClelland gives us for illustrating the model is the following: the category is members of two gangs, the Jets and the Sharks. The model’s knowledge of these individuals is represented in a node network: each node is considered as a simple processing device which continuously accumulates excitatory and inhibitory inputs from other nodes and in response adjusts its output to other nodes, much as the neuron’s adjustments of the rate of firing. There are two kinds of nodes: instance nodes, which are the nodes for each of the individuals (which are represented in the center of the figure 13), and property nodes, which are the nodes for each of the properties or attributes these individuals can have. Both instance and property nodes are arranged into thematic groups or cohorts of mutually exclusive values.

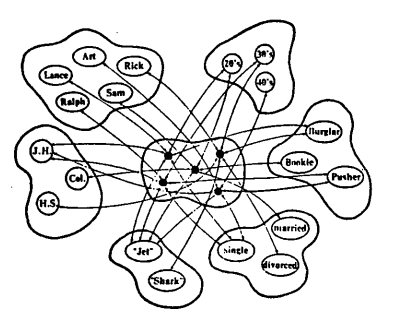


Fig. 13: The representation of several of the individuals of two gags: the Jets and the Sharks[[141]](#footnote-141).

The system’s knowledge of an individual consists simply of an instance node and a set of bi-directional excitatory links between it and the nodes for the properties that the individual is known to have. For example, the system’s representation of Lance is an instance node with mutual excitatory connections to the name node “Lance”, the gang membership node “Jet”, the age node “20’s”, the education node “Junior High” the marital status node “married”, and the occupation node “burglar”.

The system is queried by presenting it with a probe. For example, to find out about the properties of the Jets, the system can be probed by activating the property node "Member of Jets”. Before a probe is presented, each node is assumed to be at rest, with an activation value below 0. Probe presentation causes an excitatory input to be applied to each node specified in the “Jet” node. As the instance nodes of Jets node become active, they send excitation to the nodes for their properties, which drives the activation of these nodes above 0, into the “active range”. Active nodes in turn send excitatory signals to the instance nodes they are linked to and send inhibitory signals to the other nodes in the same cohort. These signals are graded, and their strength is proportional to the source node’s activation. As processing continues, the instance nodes that become active again excite the property nodes to which they are connected, inhibiting all the other instance nodes. The excitation and inhibition processes are allowed to go to equilibrium. At this point, the system has generally activated property nodes which constitute the system’s response to the probe “typical properties of the Jets”. If all of the active instance nodes "agree" on a single property, the node for that property will tend to be strongly activated, and the system would simply report that value. On the other hand, if they all specify different values within the same cohort, many values will become partially activated; the system therefore could either list the set of possibilities or make a probabilistic choice from among the alternatives.

In conclusion, this model proposed by McClelland can:

a- retrieve what it knows about specific instances;

b- induce generalizations about the shared properties of specified subsets of familiar objects;

c- fill in plausible default values for unknown properties of the retrieved individuals.

The most important difference from the symbolic models is that this model permits inducing generalizations from stored representations of specific objects or events. This conception avoids two problems presented by models which incorporate the idea of an explicit storage of generalizations: on one hand, how generalizations which are stored explicitly were obtained in the first place, a lack of explanation that we had noted about general elaborations and schemas in Anderson’s model, and on the other hand, how memory can contain explicit representations that anticipate all of the possible generalizations we might ever wish to make. Therefore, according to this model, not all the generalizations are necessarily stored as such; at least some of them are the product of an induction from particular stored information.

However, as McClelland himself recognized, the model is highly simplified and it is far from capturing the varieties of structure of different kinds of information that can be stored. Furthermore, it isn’t too different from symbolic models which incorporate the idea of spreading activation, like the ACT theory. McClelland’s model has all the characteristics we mentioned with regard to the symbolic models: a localist model represented by a network of nodes and concerned with the accuracy of the memory representation. The only difference consists in the nature of the trace: if in Anderson’s model the generalizations are encoded in a network of cognitive units, in McClelland’s model the mental general content is not stored in memory but is rather the product of an inference, so that it does not differ from the mental processes that give rise to it. We could ask, however, what happens after a generalization is induced for the first time, because it doesn’t seem plausible that it would be just a simple epiphenomen that would not have any impact on the specific stored information that originated it. Let’s imagine an example: according to this model we would arrive at the generalization “Paris is generally cloudy” from the spreading activation between two property nodes linked to the instance node “Paris”, one representing specific dates and another different types of weather[[142]](#footnote-142). Once we obtain this generalization, will it be stored or not? If it is not stored, how could we explain that sometimes we only remember the general information and not the specific information? Furthermore: memory would need a great capacity to store all the specific data indispensable to induce generalizations. If generalizations are stored once obtained, the model doesn’t explain how this occurs. As we can see, the model is too simple to give an answer to all these questions, and also to explain all memory phenomena.

*McClelland and Rumelhart’s distributed connectionist model: general/specific information*

Some years later, McClelland, together with Rumelhart, proposed a model that tries to give another answer to the specific/general information problem[[143]](#footnote-143). Their viewpoint is, in a certain way, the opposite of what they call the *enumeration of specific experiences*, which includes McClelland’s own model of 1981, and which can be summarized as the view that in memory the traces stored are traces of specific events and specific exemplars of the concepts and rules, and generalizations emerge from the superposition of them made by specific mechanisms that search the same occurrence of the concept, event, etc. In general terms, the criticism made to this consists of two points that we stressed before: this conception requires not only an unlimited storage capacity but also many mechanisms for searching what it seems an unlimited mass of data.

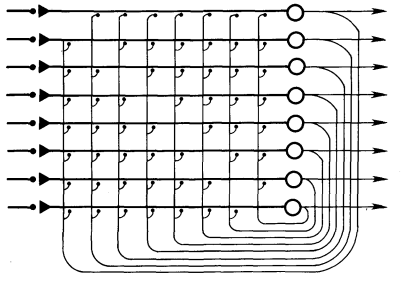
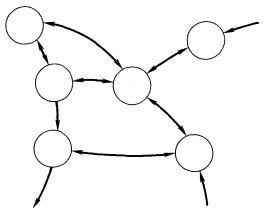
Their distributed connectionist memory model is “a model of the internal structure of some components of information processing, in particular those concerned with the retrieval and use of prior experience”, without however specifying “how these acts of retrieval and use are planned, sequenced, and organized into coherent patterns of behavior” neither having the intention of being “a complete theory of human information processing and memory”[[144]](#footnote-144). But even if their model is highly simplified, it aims to be an alternative to all the existing models and not a simple physiological implementation of existing models of cognitive processes, like Anderson and other authors aim to do, because its basic assumptions about cognition produce a totally different description about the structure of cognitive processes, as we’ll see later.

The most important difference from McClelland (1981) is that in their distributed model, even if the traces are traces of individual experiences, their superposition is not the product of a mechanism at retrieval but occurs at the time of storage, resulting automatically in abstraction: “we do not keep each trace in a separate place, but rather we superimpose them so that what the memory contains is a composite”[[145]](#footnote-145).

The basic assumptions of this model are the following:

- The smallest and basic units of the system are processing units that interconnect between each other. Because it is a distributed model, as we’ve already mentioned, a basic unit does not necessarily correspond to a particular representational primitive; a color, for example, which is generally considered as a simple and primitive property, could be a pattern of activation over a collection of units. On the other hand, because they take on activation values which range from -1 to +1, they communicate between each other “by sending signals modulates by weights associated with the connections between the units.”[[146]](#footnote-146)

- Units that are strongly connected form a module, which receives inputs and sends outputs from other modules. The state of each module represents a synthesis of the states of all of the modules from which receives input. Because some of these modules can be closer to the sensory organs and others placed at the abstract level, each module combines a number of different sources of information.

A simple information processing module of 8 A simplified diagram showing several modules andprocessing units: each one receives inputs interconnections among them[[147]](#footnote-147)

from other modules, sends outputs to other units

and has a modifiable connection to all other units

in the same module[[148]](#footnote-148)

- A pattern of activation over the units in some subset of the modules constitutes a mental state; so different mental states are simply alternative patterns of activation over the modules. A pattern of activation is the same as another if the same units are involved.

- Patterns of activation leave changes in the connection strengths between the units in the modules, so traces are not a copy of the learned pattern but changes in these strengths or weights of the units, and this is what is actually stored in memory. A memory trace is distributed over many different connections, and each connection participates in many different memory traces; this is why traces of different mental states are superimposed in the same set of weights.

- Retrieval occurs when a part of a known pattern presented for processing causes the rest of the pattern to be reinstated. This kind of fragment of the original state can be a cue originating sensory input or arising from the results of previous retrieval operations, for example.

The model described by McClelland and Rumelhart, as the authors themselves admit, involves two oversimplications: it only analyses the operations that take place within a single module, and it supposes that input from outside the module is a fixed pattern, ignoring first, that modules are in continuous interaction, and second, that external input patterns evolve in time and in general are affected by feedback from the module under study.

In order to explain the encoding of the stimulus patter, the authors divide the time of the processing cycle in ticks, which in turn are divided into two phases: in the first phase, each unit determines its net input, whereas in the second phase, the activations of the units are updated.

In the first phase, the net input of each unit is determined by the sum of the external input to the unit and the total activation of all of the units:

ni = ii + ei

where *ni* = net input of unit *i*; *ei* = the lumped external input to unit *i*; and *ii* the total input to unit *i* that are internal to the module and that is defined as follows:



which consists on the sum of all its separate inputs, where the input to unit *i* from another unit, called for example *j* (ilj), is the result of the activation of the unit *j* (*aj*) multiplied by the weight constant modulating the effect of unit *j* on unit *i*:

iij = ajwij

In the second phase, the phase of updating, two things can happen: if the net input is positive, the activation of the unit is incremented by an amount proportional to the distance left to the ceiling activation level of +1; if the net input is negative, the activation is decremented by an amount proportional to the distance left to the floor activation level of -1:

If n1 > 0 åi = En1 (1– ai) – Dai

If n1 < 0 åi = Eni [a1 – (–1)] – Dai

where *E* represents the rate of excitation; *D* a general decay factor which tends to pull the activation of the unit back toward the resting level of 0; *ai* the activation of the unit *i* at the end of the previous cycle and *åi* is the new value of the activation of unit *i* for the next cycle.

Given a fixed set of inputs to a particular unit, the activation level will continue to increase or decrease until it reaches the point where the incremental effects of the input are balanced by the decay (even if in practice, as McClelland and Rumelhart explain, it seems a little more complicated because each unit activation changes the input of the other). At this point, when weight adjustment takes place, a stable pattern of activation is achieved, and the memory trace of it will consist on the set of changes in the entire set of weights in the module.

In order to explain the process of retrieval, McClelland and Rumelhart introduce what he calls the “delta rule”. The delta rule is a weight modification that allows each part of a pattern to reconstruct the rest of the pattern, which is absolutely necessary to allow retrieval. That is, the internal connections among the units in the module have to be set up in such a way that “when part of the pattern is presented, activating some of the units of the module, the internal connections will lead the active units to tend to reproduce the rest”[[149]](#footnote-149), which means that the internal input to each unit needs to have the same effect on the unit that the external input has on the unit. The difference is measured through the following formula:

∆i = ei – ii

As Anderson explains, “in determining the activation value of the unit, we added the external input together with the internal input. Now, in adjusting the weights, we are taking the difference between these two terms”[[150]](#footnote-150). If the difference is zero, the weights do not need to be changed; if the difference is positive or negative they do: in the first case, the internal input is not activating the unit enough to match the external input to the unit, whereas in the second case it is activating the unit too much. So, in these last two cases, the weights must to be adjusted, and the formula is the following one:

ẘij = *S*∆iaj

where ẘij is the change in the weight to unit i from unit j, and *S* a global strength parameter which regulates the overall magnitude of the adjustments of the weights. Let’s quote one example mentioned by Anderson to see how it works: “consider the case in which ∆i is positive and ajis positive. In this case, the value of ∆i, tells us that unit *i* is not receiving enough excitatory input, and the value of ajtells us that unit *j* has positive activation. In this case, the delta rule will increase the weight from *j* to *i*. The result will be that the next time unity has a positive activation, its excitatory effect on unit *i* will be increased, thereby reducing ∆i”[[151]](#footnote-151). This adjustment, however, is not guaranteed, and it is only possible if the patterns conform to the linear predictability constraint, which establishes that “over the entire set of patterns, the external input to each unit must be predictable from a linear combination of the activations of every other unit.”[[152]](#footnote-152) That is why it is very important the encoding of the stimulus pattern: even if each dimension or aspect of a stimulus cannot be encoded separately, what McClelland and Rumelhart call *a* *context sensitive* encoding is necessary, such that the representation of each aspect is colored by other aspects.

As far as the decay of the trace is concerned, the authors recognized that it is a subject that has to be examined further, but they basically assume that at first the trace decay rapidly but the remaining portion becomes with time more resistant to decay.

The ability to store various patterns in the same set of units is known as *superpositioning*. This superpositioning can be between:

- Distinct patterns of activation, without one interfering with the others. When a new pattern absolutely dissimilar to the other patterns stored in the module is presented to it, the activation of the module simply reflects the direct effects of the input, as they would be in a module with 0 weights. As the module enhances the response to the new pattern, the response to the previous one is not affected; the only difference is that in a module which only stores one pattern the whole pattern can be retrieved by the activation of a single unit, whereas a module that stores several patterns needs the subpatterns of units in order to retrieve a single pattern;

- Patterns of activation that have some common structure. In this case, delta rule is used to extract this structure and throw away random variation, as in the case of learning a prototype from exemplars. After that, exemplars and prototype coexist in the same single memory module. For example, let’s imagine a boy that sees dogs and knows two dogs: the dog next door called Rover and his grandma’s dog named Fido. Each dog that he sees, included Fido and Rover, is a particular generated distortion of the dog prototype visual pattern. After learning, the model is capable of retrieving the visual pattern corresponding to the associate name given as input, as for example, “when given the Dog name pattern as input, it retrieves the prototype visual pattern for dog by only a single element”[[153]](#footnote-153); and inversely, it can also retrieve the appropriate name from a visual pattern, even if it differs from the visual pattern of dog only by a single element. The model can also complete a visual pattern given part of it: “if the part corresponds to the prototype, then that is what is completed, but if it corresponds to one of the repeated exemplars, that exemplar is completed.”[[154]](#footnote-154)

- Different prototypes, for example, when the child learns the concepts of dog, cat and bird. The more dissimilar are the patterns, the less interference there is among them; the more similar they are, the greater is the risk of confusion during learning, even if with time the delta rule finds a set of connection strengths that minimizes it.

I will not give any more details, due to their complexity. To summarize, the model:

a. captures the structure inherent in a set of different patterns creating a prototype pattern without ever being exposed to it, which is why it’s not a simple categorizer or a prototyping device;

b. retrieves accurate competitions of similar patterns through the delta rule;

c. allows the coexistence of representations of prototypes with representations of particular exemplars in the same module. Clearly, this is its most important difference from other models.

Another important point that differentiates this distributed connectionist model from the others is the conception of the influence of the repetition of a unit in the prior representation of that unit. In this model, each time a stimulus is processed, it contributes to the composite and superimposed memory representation, “giving rise to a slightly different memory trace: either because the item itself is different or because it occurs in a different context that conditions its representation.”[[155]](#footnote-155) Therefore, there isn’t a separate trace for each time the stimulus is processed, even if the composite trace also preserves the characteristics of particular experiences (at least until they are overridden by cancelling characteristics of other traces). In this aspect, this theory avoids on the one hand the encoding of multiple specific experiences, each one capturing aspects of the specific context in which it occurs, as it happens on some views, such as that of Jacoby[[156]](#footnote-156); and on the other hand, the multiplication of context-specific detectors that are frequent in so-called logogen theories which introduce specialized recognition units for each specific word[[157]](#footnote-157), and which requires complex acquisition mechanisms. So, as Anderson explains, this distributed connectionism model captures the generalizations economically and through a simple connection strength modulation scheme, explaining the abstraction and the preservation of information about specific stimuli through the same basic learning mechanism, which permits minimizing the size of the network. What’s more, it allows modification of the memory trace after a new stimulus is processed, at least with respect to prototype representation.

Consequently, this distributed connectionist model has some advantages relative to the other models, even if we have to recognize that these advantages come more from the idea of distributed connectionism than from this specific presentation of it, which is very simplified and only analyses the case of the encoding and retrieval of general concepts and the specific instantiations of them. The distributed connectionist framework:

- overcomes the symbolic models that reduce memory to a linguistic phenomena; it therefore seems to be a more appropriate common denominator for the great variety of types and manifestations that most would consider to fall under the rubric of memory;

- facilitates the communication between memory models belonging to different levels of analysis like neural and functional models of memory;

- seems to be the most economical model, not only because of its explanation of the preservation of general and specific information but also because it explains through the same mechanisms analogy, generalization and errors, like for example cross-talk[[158]](#footnote-158).

One important problem that distributed connectionism seems to present is the problem of the representation of time: how does connectionism capture the linear succession of temporal event? Even if some authors, like Elman[[159]](#footnote-159), propose a model which tries to capture something of the flavor of temporal order, this still poses a problem to connectionism. In this sense, the problem underlying the difficulty of representing temporal order seems to be the same problems presented in order to account for the syntactic and semantic rules, as we will see later.

We can affirm that the major idea that differentiates distributed connectionism from rival proposals is a more holistic and dynamic conception of the memory system[[160]](#footnote-160), a conception which implies that:

(1) The system is interdependent: activation in one part of the system is nearly always affected by activations in other parts. The current state of knowledge of the system determines encoding of information as well as retrieval; for example, what is most salient in a memory can change over time depending on information elsewhere in the system, making in some sense true the idea that remembering an episode may never be quite the same twice;

(2) The system is in a continuous state of change: The effect of one unit on another is not necessarily constant, because experience can change the weights and the connections, thus altering the knowledge of the system. In some cases this alteration is a form of learning, which constitutes “the most important and interesting aspect of connectionist models: the ability to learn”[[161]](#footnote-161). Even if in McClelland and Rumelhart’s model it seems possible to reach identical states many times, this is due to the simplification of the model and it’s much less likely to happen in a well-developed distributed connectionist model (as well as in real life), where identical states of activity would be the exception rather than the rule.

Furthermore, distributed connectionism seems to be the only model that could be compatible with the idea that changes can come not only from changes in the system (as the result for example of the encoding of new information which seems to be a continuous activity performed by the system) but also from the simple retrieval of information, as the new studies on reconsolidation seems to prove. I will go into this point in more depth. The consolidation hypothesis, which constituted a dogma for over a century, says that when we encode some experience the synapses in the brain involved in the encoding are strengthened until they are stabilized in the form of new synaptic connections that then make the memory persist; retrieval is thought as reactivation of these connections. However, two types of studies contradict this idea:

- Studies principally made by Ledoux’s laboratory research[[162]](#footnote-162) show that memories undergo a process of reconsolidation, which involves the creation in long-term memory of new synaptic connections via protein synthesis, every time they are recalled, which suggests that memories aren’t simply restored each time they are retrieved.

- Studies made by Nadel and Moscovitch[[163]](#footnote-163) show that the hippocampus, whose intervention seemed no longer to be needed after sufficient repetition over time when neocortical sides became directly connected and the trace seemed consolidated, intervenes each time the trace is retrieved, at least as far as episodic memory is concerned[[164]](#footnote-164). The multiple trace theory (MTT) that they propose supposes that each time it is retrieved from the neocortex it re-enters into hippocampal circuits and is re-encoded, so retrieval would constitute a new encoding event.

So this conception of “reconsolidation” implies that retrieval is not a passive mechanism and does not clearly leave the experience retrieved the same, even from a molecular perspective. Even if consolidation is also an essential part of the whole memory process, “each time a memory is retrieved, it is integrated into ongoing perceptual and emotional experiences and becomes part of a new memory”[[165]](#footnote-165). For example, Gardiner[[166]](#footnote-166) questions the linear encoding – storage – retrieval view, which loses sight of the fact that any retrieval event is also an encoding event: encoding and retrieval involve continually interchangeable, constructive and reconstructive processes and complex interactions.

Consequently, the concept of reconsolidation “reflects the fact that is important for traces of prior experience to be left open to modification by future events. Such traces are not only subject to passive processes such decay and the active but indirect process of interference. They are also subject to meaningful transformation as a function of new experience”[[167]](#footnote-167). It is difficult to think that a symbolic account, even a well developed account like Anderson’s model, could be compatible with what is implied in the notion of “reconsolidation”: retrieval is conceived as the reactivation of the nodes of the network that are involved in the trace, and it’s a requirement of recall that all the elements are reactivated. Even more, Anderson’s model does not even conceive the possibility of a modification which does not have a quantitative nature: the only thing that happens when a trace is retrieved is the increase of its long-term strength by one unit. Therefore, only a distributed connectionist model will be able to incorporate the notion of reconsolidation and the idea initially promoted by Tulving[[168]](#footnote-168) that retrieval occurs as a result of integration of incoming change in environmental information with the memory network driven by that information.

The only problems that arise from this conception of retrieval and reconsolidation are, on one side, the identity of the memory: does retrieval lead to modification of the original memory or to the formation of a new memory[[169]](#footnote-169)?, and on the other hand, the distinction between memory and imagination: how do we distinguish between a memory of an event that has happened and an event that was just imagined or just planned but not executed? I will return to this questions in the conclusion.

(3) As I noted above, this model is the only one which explains in the most economical way the preservation and relation between general and specific information. Moreover, it suggests that semantic memory is not really a separate system but only the residue of the superposition of episodic traces: “Consider, for example, representation of a proposition encountered in several different contexts, and assume for the moment that the context and content are represented in separate parts of the same module. Over repeated experience with the same proposition in different contexts, the proposition will remain in the interconnections of the units in the proposition submodule, but the particular associations to particular contexts will wash out. However, material that is only encountered in one particular context will tend to be somewhat contextually bound”[[170]](#footnote-170). This is also compatible with Smith’s notion of the “degree of insularity” of memories[[171]](#footnote-171), which ranges from insular mental memories bounded to unique contexts which are inaccessible when context cues are not provided (which is generally the case with episodic memories), to completely decontextualized memories that need no contextual referents because they have been retrieved and used in many varied contexts and can be accessed regardless retrieval context (as happens in general with semantic memories). This conception however would have to be compared principally with Tulving’s empirical studies about the existence of two declarative memory systems that would be in certain way independent from each other[[172]](#footnote-172). Anyway, the idea of semantic memory being the result of episodic memories is very suggestive. I will also come back to this point in the conclusion.

Therefore, for the moment this distributed connectionist model can account simultaneously for persistence and transformation, as well as for equivalence and context sensitivity. It is necessary to remark again that McClelland and Rumelhart’s model is a very simple one: it only explains general concepts and their particular instantiations. This seems to be a constant in the distributed connectionist models that have been developed so far: as we’ve already mentined, they only explain “lower” cognitive processes: recognition of words, the past of verbs, the coexistence of general/specific memories, etc[[173]](#footnote-173). What is very attractive in connectionism is more its basic assumptions than the specific models that have been created on their basis. Regardless, these basic assumptions are extremely interesting and lead to these other three suggestions about memory and cognition:

(A) In spatial analogies as well as in symbolic models the memory was the same as the memory trace; and moreover, the trace continues to conserve its individuality and its boundaries in connection with other traces. Only in a distributed connectionist model we can conceive, on one hand, the memory trace not as a kind of entity, but as a mental processing, as a change in the system; and on the other hand, the memory as a kind of emergent entity that first, is different from the trace and second, does not have an epiphenomenal character (as we shall see shortly). Concerning the first point, the idea is to conceive the memory trace not as a kind of material entity but as a change, that is, as a “relationship (difference) between two things that are physical objects, the brain at time 1 ('immediately before' the experience) and time 2 ('immediately after' the experience), but the relationship itself is not a physical object”[[174]](#footnote-174). Concerning the memory, on one hand, it is different from the memory trace because the later constitutes only one condition of possibility of the former, which would not be possible if the memory trace would not interact with retrieval cues and information derived from particular environment conditions: “memory does not exist independently of its being retrieved”[[175]](#footnote-175); on the other hand, the memory event that emerges from this interaction is not a simple epiphenomenon because it has a causal power over the trace: the act of remembering may involve and give rise to patterns of neural activity that, in turn, change the traces that have themselves participated in that memory’s expression (which would explain the reconsolidation phenomenon). This dynamic dialectic between the memory event and the memory traces can’t be conceived in a symbolic model where the memory event can only increase the strength of the trace. What’s more, even if it’s not implied by distributed connectionism, only by assuming this kind of model can we question the idea that reality is already segmented in episodes, events or facts with delimited boundaries (and maybe a start, a development and an end) that are encoded as wholes in separate units (or connections between units).

(B) Only distributed connectionism and not symbolism seems compatible with an external symbolic approach. Fodor and Pylyshyn[[176]](#footnote-176) have criticized connectionism on the grounds that it can’t account for some features that cognition manifests, like the fact that representations are built according to compositional and syntactic rules. It is true that connectionism lacks an internal representational system with composition syntax and semantics. Nonetheless, as Bechtel explains[[177]](#footnote-177), a possible answer consists in explaining that inside the cognitive system there are not internal representations of these external symbols but an ability to extract information from them and to produce symbol strings which adhere to the syntactic rules that characterize properly formed strings. Connectionism would allow us to conceive cognition as a network that knows how to extract information of the environment and has adapted to a linguistic environment, and to conceive language as a social device for categorizing information about the environment that even if it allows us to acquire information that wouldn’t be available to us otherwise, does not constitute the only way of categorization; this would also explain the fact that before learning language we have already learned to categorize a our environment and we can govern our actions in accord with those categorizations. In this sense, connectionism would clearly be compatible with Vygotsky’s old idea of language as an external tool that the child appropriates and internalizes through the social interaction and incompatible with Chomsky’s idea of an universal grammar; that is, with a social approach of acquisition of language and not with an nativist point of view. In any case, as the same Bechtel recognizes, explaining how a system can extract from language syntactic and compositional rules constitutes a challenge for connectionism.

(C) Distributed connectionism is also the only conception that would be compatible with a situated approach to memory, which rejects any kind of localist description of memory and considers that memory operates in an extended system of cultural and personal habits which shapes our inner representational system and this in two ways “by treating inner representations and traces as often incomplete, partial, and context-sensitive, to be reconstructed rather than reproduced, and by widening the representational realm outside the organismic boundary”[[178]](#footnote-178). I will not go deeply into this question, which clearly demands an analysis of the plausibility of the hypothesis of the extended mind; I only want to add two things: first, that cases like collaborative recall in small groups, families or couples about shared experiences as well as group’s production of a shared and social memory (as happens with survivors and relatives of death or disappeared people of the last dictatorship in Argentina) are difficult to explain outside a situated approach to memory; second, that only in a situated analysis of memory can we move beyond the correspondence idea of truth in memory, which is only a simple consequence of the causal theory of memory, which establishes a correspondence between a “past event”, a trace and the memory of that trace, and start to rethink memory more from a functionalist and pragmatic point of view, enlarging the notion of “accuracy” to for example the notion of “good remembering”[[179]](#footnote-179).

In conclusion, the development of these different memory models shows that:

(a) Some assumptions blindly accepted by the causal theory of memory are already questioned in symbolic models, such as the biunivocal correspondence between a past representation and a memory trace, the existence of a categorical boundary between episodic and semantic memories, etc.;

(b) Distributed connectionism offers a real possibility of conceiving a new philosophical theory of memory which could explain more adequately the richness of real-life memory phenomena, even if the development of this model constitutes a difficult task.

**CONCLUSION**

As I have tried to show, the causal theory of memory, which is the only theory developed so far that at first view seems more plausible and that could be integrated with psychological explanations and investigations of memory, shows some conceptual and ontological problems that go beyond the internal inconsistencies that each version can present. On one hand, the memory phenomenon analyzed is very limited: in general it is reduced to the conscious act of remembering expressed in a propositional format: the idea of an agent who can control his memory reports, in contrast to a passive subject who merely retrieve his encoded memories, is not even considered. Furthermore, the representational function of memory is the only function taken into account; and all the diachronic changes that could entail a dynamic conception of memory are ignored, as is the consideration of context in a broad sense of the term. On the other hand, causal theories make some implicit assumptions that are obviously questionable (e.g., the difference between facts and events) and leave some of their key concepts unexplained, like the nature of the causality, the isomorphism between the memory trace and the memory event, etc. Nonetheless, the principal point that we criticized is the idea of a correspondence between a past event or representation of it, the correspondent memory trace and the memory representation of it.

Psychological models of memory, with the exception of the spatial analogy, question some of these assumptions and suggest that memory information is organized in a way other than that implied by the causal theory. But it is distributed connectionism that actually challenges the causal theory of memory, not only through some of its basic postulates, like the superpositional conception of memory traces, their mutual influence and constant state of change, but also because it is compatible with a more situated approach to memory phenomena as well as with the idea that memory in reality accomplishes a variety of functions that go beyond the representational one.

The explicit and implicit ideas provided by distributed connectionism, together with the previous criticisms that we made to of causal theory, clearly suggest the possibility of developing another kind of philosophical theory of memory, different from existing theories. This task, nonetheless, implies:

(a) A much deeper study of connectionism principles and memory models, as well as of the rival candidate, symbolism, plus the evaluation of these models in the light of psychological and neurophysiological empirical research, which unfortunately I have not dealt with in this work.

(b) An analysis of the metaphysical nature of events and facts, in order to see if we should decide in favor of a realist or a non-realist account. It is also clear that this decision can’t be taken without an examination of perception studies (and not exclusively philosophical ones) which would be crucial for determining what is encoded in memory, and studies devoted to language, which, I suspect, plays an important role in the segmentation of reality.

(c) A clarification of the notion of causality, because we wouldn’t like to deny any kind entail the complete disappearance of boundaries between memory and imagination. And even if I suspect that my future work will question the existence of absolute boundaries between these two capacities, the notion of causality can’t be completely erase from the theory of memory without erasing the memory phenomenon itself. But this causality, as we’ve already showed, can’t consist in the naïve conception adopted by the causal accounts of memory, nor can it be explained in counterfactual terms. If we think of the reconsolidation concept, of the interdependence between all the memory system, and of the deciding influence of context, it’s almost sure that more than one notion of causality will need to be invoked in order to explain memory.

(d) Because I do not want to reduce the study of memory exclusively to cases of remembering, it’s essential to analyze how memory intervenes not only in imagination, but also in perception, reasoning and future projection, for example. In this respect, we could quote the interesting articles of Atance & O’Neill (2001) and Bucker & Carroll (2006), the first concerning the relationship between episodic memory and episodic future thinking, the second concerning the relationship between episodic memory and all kinds of self-projection, such as navigation and theory of mind, relations that would also have to be explored from a philosophical perspective, in particular to overcome the tendency to suppose that representation is the only function of memory. I will also mention the necessity of combining the notion of metamemory with that of memory to some extent; as I have already remarked, it is implausible to suppose that a subject who consciously remembers something is unable to exert any kind of control over his memory reports.

(e) Finally, in order to integrate a situated conception of memory into a philosophical account, it will also be indispensable to take into consideration what is known as “everyday memory research” as well as studies in social and collective memory.

**BIBLIOGRAPHY**

Anderson, J. R. (1983a). A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior*, 22, pp. 261-295.

Anderson, J. R. (1983b). *The Architecture of Cognition*. New Jersey, Lawrence Erlbaum Associates.

Anscombe, G. E. M. (1963). The Reality of the Past. In M. Black (Ed.), *Philosophical Analysis: A Collection of Essays*. London, Cornell University Press, pp. 36-56.

Bergson, H. (1896). *Matière et Mémoire*. Paris, Les Presses universitaires de France, 1965.

Bernecker, S. (2008). *The Metaphysics of Memory*. Dordrecht, Springer.

Bechtel, W. (1993). The Case for Conectionism. *Philosophical Studies*, 71, 119-154.

Borges, J. L. (1944). Funes el Memorioso. In J. L. Borges, *Ficciones*. Buenos Aires, Editorial Sur.

Broad, C. D. (1925). *The Mind and Its Place in Nature*. London, Kegan Paul, Trench, Trubner & Co.

Campbell, S. (2006). Our Faithfulness to the Past: Reconstructing Memory Value. *Philosophical Psychology*, 19(3), pp. 361-380.

Casati, R. & Varzi, A. (2006). Events. In *Stanford Encyclopedia of Philosophy*.

Collins, A. R & Hay, D. C. (1994). Connectionism and memory. In P. E. Morris & M. Gruneberg (Eds.), *Theoretical Aspects of Memory*, Londres, Routledge, pp. 195-237.

Collins, A. R. & Quillian, M. R. (1969). Retrieval Time from Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, 8, pp. 240-247.

Collins, A. R. & Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, 82(6), pp. 407-428.

Deutscher, M. (1969). A Causal Account of Inferring. In R. Brown and C. D. Collins (Eds.), *Contemporary Philosophy in Australia*, London, George Allen & Unwin, pp. 97-118.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MIT Press.

Dretske, F. (2000). The Mind’s Awareness of Itself. In F. Dretske, *Perception, Knowledge and Belief: Selected Essays*. New York, Cambridge University Press.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, 14, pp. 179-211.

Fernandez, J. (2008). Memory, Past and the Self. *Synthese*, 160, pp. 103-121.

Floridi, L. (2005). Semantic Conceptions of Information. In *Stanford Encyclopedia of Philosophy*.

Fodor, J. A. & Pylyshyn, Z.W. (1988). Connectionism and Cognitive Architecture: A Critical Analysis. [*Cognition*,](http://www.ncbi.nlm.nih.gov/pubmed/2302942) 28, pp. 3-71.

Gardiner, J. M. (2007). Retrieval: On Its Essence and Related Concepts. . In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*, New York, Oxford University Press, pp. 171-176.

Jacoby, L. L. (1983a). Perceptual Enhancement: Persistent Effects of an Experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,9, pp. 21-38.

Jacoby, L. L. (1983b). Remembering the Data: Analyzing Interaction Processes in Reading. *Journal of Verbal Learning and Verbal Behavior*, 22, pp.485-508.

Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach*. Institute for Cognitive Science Report 8604, University of California, San Diego.

Kierkegaard, S. (1843). *La Reprise*. Paris, GF Flammarion, 1990.

Koriat, A. & Goldsmith, M. (1996). Memory Metaphors and the Real-life/Laboratory Controversy: Correspondence versus Storehouse Conceptions of Memory. *Behavioral and Brain Sciences*, 19, pp. 167-188.

Lackey, J. (2005). Memory as a Generative Epistemic Source. *Philosophy and Phenomenological Research*, 70 (3), 636-658.

Lebiere, C. & Anderson, J. R. (1993). A Connectionist Implementation of the ACT-R Production System. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, pp. 635-640.

LeDoux Consolidation: Challenging the Traditional View. In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York, Oxford University Press, pp. 171-176.

Loeb, L. E. (1974). Causal Theories and Causal Overdetermination. *The Journal of Philosophy*, 71 (15), pp. 525-544.

Mackie, J. (1974). *The Cement of the Universe*. Oxford, Oxford University Press.

Malcolm, N. (1963). A Definition of Factual Memory. In N. Malcolm, *Knowledge and Certainty*. London, Cornell University Press, pp. 222-240.

Malcolm, N. (1970). Memory and Representation. *Noûs*, 4(1), pp. 59-70.

Malcolm, N. (1977). *Memory and Mind*. London, Cornell University Press.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, Freeman.

Martin, C. B. & Deutscher, M. (1966). Remembering. *The Philosophical Review*, 75(2), pp. 161-196.

Matthen, M. (2010). Is Memory Preservation? *Philosophical Studies*, 148(3), pp. 3-14.

McClelland, J. L. (1981). Retrieving General and Specific Information from Stored Knowledge of Specifics. *Proceedings of the Third Annual Conference of the Cognitive Science Society*, Berkeley.

McClelland, J. L. & Rumelhart, D. E. (1985). Distributed Memory and the Representation of General and Specific Information. *Journal of Experimental Psychology: General*, 114(2), pp. 159-188.

McKay, T. & Nelson, M. (2005). Propositional Attitude Reports. In *Stanford Encyclopedia of Philosophy*.

Michaelian, K. (2010). Generative Memory. *Philosophical Psychology* (forthcoming).

Morton, J. (1979). Facilitation in Word Recognition: Experiments Causing Change in the Logogen Model. In P. A. Kohlers, M. E. Wrolstal & H. Bouma (Eds.), *Processing Visible Language I*.New York, Plenum.

Moscovitch, M. (2007). Memory: Why the Engram is Elusive? In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York, Oxford University Press, pp. 17-22.

Nadel, L. (2007). Consolidation. The Demise of the Fixed Trace. In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York, Oxford University Press, pp. 177-182.

Nadel, L. & Moscovitch, M. (1997). Memory Consolidation, Retrograde Amnesia and the Hippocampal Complex. *Current Opinion in Neurobiology*, 7, 217-227.

Naylor, A. (1971). On “Remebering” an Unreal Past. *Analysis*, 26, pp. 29-41.

Neisser (1988). What is Ordinary Memory the Memory of? In E. Winograd & U. Neisser, *Remembering Reconsidered : Ecological and Traditional Approaches to the Study of Memory*. New York, Cambridge University Press, pp. 356-373.

Quillian, M. (1962). A Revised Design for an Understanding Machine. *Mechanical Translation*, 7(1), pp. 17-29.

Quillian, M. (1968). Semantic Memory. In M. Minsky (Ed.), *Semantic Information Processing*. Cambridge, MIT Press, pp. 227-270.

Pylyshyn, Z. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*, New York, MIT Press.

Rips, L. J., Shoben, E. J. & Smith E. E. (1973). Semantic Distance and the Verification of Semantic Relations. *Journal of Verbal Learning and Verbal Behavior*, 12(1), pp. 1-20.

Raaijmakers, J. G. W. & Shiffrin, R. M. (1981). Search of Associative Memory. *Psychological Review*, 88(2), pp. 93-134.

Robinson, H. (1994). *Perception*. London, Bantam.

Roediger, H. I. (1980). Memory Metaphors in cognitive Psychology. *Memory & Cognition*, 8(3), pp. 231-246.

Rosch, E. (1975). Cognitive Representation of Semantic Categories. *Journal of Experimental Psychology*, 104, pp. 192-233.

Russell, B. (1921). *The Analysis of Mind*. London, George Allen & Unwin.

Sara, S. J. (2007). Consolidation: From hypothesis to paradigm to concept. In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York, Oxford University Press, pp. 183-190.

Schacter, D. (2007). Memory: Delineating the Core. In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York, Oxford University Press, pp. 23-28.

Schaffer, J. (2003). Overdetermining Causes. *Philosophical Studies*,114, pp.23-45.

Schumacher, J. A. (1976). Memory Unchained Again. *Analysis*, 36 (2), pp. 101-104.

Sherouse, M. A. (1979). *Memory and Knowledge: An Examination of the Epistemic Theory of Memory*. Dissertation, Ohio State University.

Shoemaker, S. (1970). Persons and Their Pasts. *American Philosophical Quarterly*, 7(4), pp. 269-285.

Silva, A. J. (2007). Consolidation: Molecular Restlessness. . In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York, Oxford University Press, pp. 167-170.

Smith, S. M. (2007). Context: A Reference of Focal Experience. . In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York, Oxford University Press, pp. 183-190.

Spear, N. E. (2007). Retrieval: Properties and effects. In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York, Oxford University Press, pp. 111-114.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, 4, pp. 411-444.

Strevens, M. (2007). Mackie Remastered. In J.K. Campbell and M. O'Rourke, *Topics in Contemporary Philosophy vol. 4: Causation and Explanation*. Cambridge, MIT Press.

Sutton, J. (2009). Remembering. In P. Robbins & M. Aydede (Eds.), Cambridge Handbook of Situated Cognition. New York, Cambridge University Press.

Thomson, D.M., & Tulving, E. (1973). Associative Encoding and Retrieval: Weak and Strong Cues. *Journal of Experimental Psychology,* 86, pp. 255-262.

Tulving, E. (1997). On the Uniqueness of Episodic Memory. In L. S. Nilson & M. J. Markowitsch (Eds*.), Cognitive Neuroscience of Memory*, pp. 11-42.

Tulving, E. (2005). Episodic Memory and Autonoesis: Uniquely Human? In H. S. Terrace & J. Metcalfe (Eds.), *The Missing Link in Cognition*. New York, Oxford University Press, pp. 4-56.

Tulving, E. (2007). Coding and Representation: Searching for a Home in the Brain. In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York, Oxford University Press, pp. 65-68.

1. For the moment, we’re going to use indistinctly “memory event” or “memory representation”, only to avoid the ambiguity of the english term “memory”. We reserve so this last word to the general capacity. [↑](#footnote-ref-1)
2. As we will see later, the nature of the memory trace differs according to the different accounts (even if it is not generally well specified). [↑](#footnote-ref-2)
3. We use the term past experience to refer to whatever the subject saw, heard, learned, etc, in the past because we consider that he has to have experienced it in some of these ways in order to being able to subsequently remember it, even if the thing remembered is not a personal experience but a fact of the world. See, for example, Ascombe (1963). [↑](#footnote-ref-3)
4. Martin & Deutscher (1966), p. 166. [↑](#footnote-ref-4)
5. Ibid., p. 174. [↑](#footnote-ref-5)
6. For Martin & Deutscher, “nobody actually remembers anything until he comes to the point of representing in some way what he has observed or experienced. We intend the vaguenesss of the phrase 'represents the past'. Already in connection with memory without belief, we have described an example in which painting was a case of remembering. At the beginning of the article we suggest that even to swim might be a form of representation.” Ibid., p. 172. [↑](#footnote-ref-6)
7. Ibid., p. 181. [↑](#footnote-ref-7)
8. Ibid., p. 179. [↑](#footnote-ref-8)
9. “A prompting is complete if the person cannot correctly give back any more in his representation of what happened in his past than was supplied by the prompting.” Ibid., p. 182. [↑](#footnote-ref-9)
10. Ibid., p. 191. [↑](#footnote-ref-10)
11. This case would refer to what psychologists call *implicit remembering*, which involves influences of prior episodes on current behavior without intentional retrieval and sometimes without conscious remembering of those prior episodes. [↑](#footnote-ref-11)
12. See Tulving (2005). [↑](#footnote-ref-12)
13. See Russell (1921), chapter IX. [↑](#footnote-ref-13)
14. Martin and Deutscher (1966), p. 166. [↑](#footnote-ref-14)
15. In order to clarify this difference, we can imagine two situations. In situation 1, a person cracks a cup (A), the crack persists and becomes bigger (B), and after a while the cup breaks (C). In the situation 2, a person breaks a cup (A), some pieces of the cup falls down into the floor (B), another person arrives without shoes and cuts the sole of her foots (C). It’s clear that in case 1, even if there is a temporal gap, the action (A) and the action (C) are causally closer than in the situation 2. In case 1 (B) is only a trace of the act (A), which is why (A) seems to be the direct cause of (C), whereas in case 2 (A) is only the cause of (B), the latter being the direct cause of (C). [↑](#footnote-ref-15)
16. Ibid., p. 185. [↑](#footnote-ref-16)
17. In an article dedicated to causality called “A causal account of inferring” (1969), Deutscher gives us two examples of causality in ordinary life that can help us to understand the difference between the two cases. For “operative for”, he invites us to imagine that “a brick is thrown according to the pre-arranged signal of a low-pitched whistle. It is partly because of the whistle that the brick is thrown. The brick then breaks a window. Thus the low-pitched whistle is, by transitivity, an operative condition *for* the window’s being broken. But it is not operative in bringing it about that the window is broken *in* the circumstance that a brick was thrown at it.” A case of “operative for and in” is the following one: “A tennis ball is thrown according to the pre-arranged signal of a high-pitched whistle. It is partly because of the whistle that the ball is thrown. The ball may not be thrown hard enough to break the window by itself. The additional effect of the sound waves from the whistle may also be needed. Thus the high-pitched whistle is operative both *for* and *in* the circumstance that the ball is thrown” (page 110). It is interesting to observe the analogy between the sound waves of the whistle, which are like a prolongation of it, and the mental trace left by an experience, which could also be viewed as a prolongation of the existence of something that has ceased to exist. [↑](#footnote-ref-17)
18. Martin & Deutscher (1966), p 190. [↑](#footnote-ref-18)
19. Malcolm (1977), chapter 5 and 9. Anyway, we are going to develop this point in at the end of this first part. [↑](#footnote-ref-19)
20. Martin & Deutscher (1966), p. 171. [↑](#footnote-ref-20)
21. We can suppose that Martin & Deutscher would also conceive of the causal relation between the public event and the initial representation in terms of operativeness, in the sense that the public event may be neither sufficient to produce a representation nor (because the same representation could be produce by other means) necessary. [↑](#footnote-ref-21)
22. Bernecker (2008), p. 21. [↑](#footnote-ref-22)
23. Martin & Deutscher (1966), p. 176. “To decide that he would not have done so is to decide that his past witnessing is causally necessary for his present account”. But immediately after that, they say in parenthesis “We shall see that a cause need not to be a necessary condition. Naturally, in such cases it is difficult to verify that someone does remember what he recounts.” [↑](#footnote-ref-23)
24. Bernecker (2008), p. 21. [↑](#footnote-ref-24)
25. Martin & Deutscher (1966), p. 185. [↑](#footnote-ref-25)
26. The term “representation” includes the two cases mentioned in the clause 2, made them reference either to the pair public events/private episodes (or episode/fact, which is less probable). See footnote 5. [↑](#footnote-ref-26)
27. Unfortunately, I did not have access to the newer book until very recently. It seems that Bernecker reformulates some points that were not well-developed in the first book and goes into some subjects in more depth, especially concerning the classification of memory, memory causation and the nature of traces. Regrettably, I must leave the analysis of his latest ideas for another work. [↑](#footnote-ref-27)
28. Two aspects that, however, will lead us to discuss almost all of the topics developed in the book. [↑](#footnote-ref-28)
29. Bernecker (2008), p. 2. [↑](#footnote-ref-29)
30. Broad (1925), p. 223. [↑](#footnote-ref-30)
31. See Squires (1992). [↑](#footnote-ref-31)
32. Malcolm (1977), pp. 66-79. [↑](#footnote-ref-32)
33. Or maybe he does not want to give any kind of justification, in order to avoid these problems (leaving aside the empirical grounds, which are absolutely absent from all Bernecker’s analysis). [↑](#footnote-ref-33)
34. See Casati & Varzi (2006). [↑](#footnote-ref-34)
35. In this sense, it is clear that Bernecker implicitly admits Dretske’s notion of a difference between awareness of properties, of objects and of facts, and also his idea of the possibility of being aware of a property without it being instantiated in or by some object. See Dretske (2000). [↑](#footnote-ref-35)
36. Bernecker (2008), p. 3. [↑](#footnote-ref-36)
37. The difference between the occurrent use and the dispositional sense is an act/potency difference: “When we say of S that he remembers that p we either say of him that he is currently engaged in propositional remembering or that he is disposed to remember that p if suitable circumstances occur”. Ibid., p. 3. It is not new; we can find it for example in Broad (1925) p. 223. [↑](#footnote-ref-37)
38. See Malcolm (1963). [↑](#footnote-ref-38)
39. The concept was first proposed by Frege in 1892. See McKay & Nelson (2005). [↑](#footnote-ref-39)
40. Bernecker (2008), p. 5. [↑](#footnote-ref-40)
41. Ibid. [↑](#footnote-ref-41)
42. Martin & Deutscher (1966), p. 189. [↑](#footnote-ref-42)
43. Bernecker (2008), p. 41. [↑](#footnote-ref-43)
44. Ibid., p. 41. [↑](#footnote-ref-44)
45. See Broad (1925), pp. 852-856. [↑](#footnote-ref-45)
46. Schumacher (1976), Squire (1992). [↑](#footnote-ref-46)
47. For example this quote from Dretske seems to illustrate well what is implicit in Bernecker’s account: “The mathematical theory of information may be an elegant device for codifying the statistical features of, and mutual dependencies between, those physical events on which communication depends, but information has to do, not with the vehicles we use to communicate, but with what we communicate by means of them. A genuine theory of information would be a theory about the content of our messages, not a theory about the form in which this content is embodied.” Dretske (1981), p. 40. This characterization, however, is common: “Philosophical approaches differ from MTC in two main respects. First, they seek to give an account of information as *semantic* content, investigating questions like 'how can something count as information?' and 'why?'” Floridi (2005). [↑](#footnote-ref-47)
48. Marr differentiates between the computation (“what is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?”), the representation and algorithm (“how can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?”) and the hardware implementation (How can the representation and algorithm be realized physically?) Marr (1982), p. 25. Pylyshyn’s description of the three levels is very similar: “there are actually two distinct levels above the physical (or neurophysiological) level –a representational or semantical level and a symbol-processing level.” Pylyshyn (1984), p. 24. [↑](#footnote-ref-48)
49. Bernecker (2008), p. 36. [↑](#footnote-ref-49)
50. Bernecker (2008), p. 154. [↑](#footnote-ref-50)
51. Shoemaker (1970). [↑](#footnote-ref-51)
52. Malcolm (1963). [↑](#footnote-ref-52)
53. Squire (1976). [↑](#footnote-ref-53)
54. Bernecker (2008), p. 26. [↑](#footnote-ref-54)
55. See, for example, Naylor (1971), Sherouse (1979). [↑](#footnote-ref-55)
56. Bernecker (2008), p. 52. [↑](#footnote-ref-56)
57. For the definition of complete prompting, see footnote 8; strict prompting has to do with the mode in which information is presented: “When he is thus verbally prompted, so that he is even told that he has experienced before what he is being informed about, we shall say that he is strictly prompted.” Martin & Deutscher, p. 183. [↑](#footnote-ref-57)
58. Loeb (1974). [↑](#footnote-ref-58)
59. See, for example, some recent articles: Schaffer (2003), Strevens (2007). [↑](#footnote-ref-59)
60. Bernecker (2005), p. 53. [↑](#footnote-ref-60)
61. Ibid., p. 54. [↑](#footnote-ref-61)
62. Bernecker quotes Mackie’s definition differently: “an insufficient but necessary part of an unnecessary but exclusively sufficient condition of some event”. His objection that this definition is too strict in order to cover all kinds of memory causation is based on the expression “exclusively sufficient condition”, which is absent from the original formulation of the “inus condition”. See Mackie (1974). [↑](#footnote-ref-62)
63. Bernecker (2005), p. 65. [↑](#footnote-ref-63)
64. Ibid., p. 66. [↑](#footnote-ref-64)
65. Ibid., p. 67. [↑](#footnote-ref-65)
66. Malcolm (1977), p. 36: “It seems to me that the only intelligible part of the half-articulate notion that memory is a form of immediate acquaintance or direct awareness lies in its rejection of the conception, more commonly favored by philosophers, that when we remember some past object the remembering is done by means of a present image, copy, or representation of the object. This negative criticism is sound. The positive aspect of the notion, the striving to say that in remembering the object of memory is 'compresent', or 'persists into the present', or 'has a mode of being', presents nothing more than vague pictures. I think it just to remark that the intelligibility of the direct awareness doctrine consists exclusively in its saying 'No!' to the representative theory of memory.” In chapter 2 of this book, Malcolm summarize some direct realist accounts of memory, like those of T. Reid, S. Alexander and W. Earle. The principle is exactly the same as Bernecker’s: the claim that in remembering a past experience we do not have a representation of it but are “direct aware” of the past experience. [↑](#footnote-ref-66)
67. Robinson (1994). [↑](#footnote-ref-67)
68. He writes “perception” and not “memory” because he’s explaining the general conception of direct realism, taking in consideration indistinctly arguments and accounts from both memory and perception. For our purpose, we could change the words “perception” for “memory” without any problem. [↑](#footnote-ref-68)
69. Bernecker (2008), p. 70. [↑](#footnote-ref-69)
70. Like the same Bernecker says. See p. 153. [↑](#footnote-ref-70)
71. I will avoid taking into consideration one idea described by Bernecker in some paragraphs, namely the following: “while the truthmaker of meta-representational memory (of a past non-factive attitude such as belief) is our mind, the truthmaker of reproductive memory is our mind and the world”, p. 140. The idea of the mind being the only truthmaker of a meta-representational memory is contradicted some paragraphs later, where he specifies that the first-order content must also be true, as we’ll see next. [↑](#footnote-ref-71)
72. Ibid. [↑](#footnote-ref-72)
73. Bernecker avoids dealing with the relation between personal identity and memory in this work. He only assumes a bodily criterion of personal identity, which, unfortunately, he considers only as an aspect of the factivity constraint on meta-representational memory. [↑](#footnote-ref-73)
74. Once more the example does not refer to a fact, but rather to an event. But we can easily generate a similar example for fact memory. [↑](#footnote-ref-74)
75. Ibid., p. 151. [↑](#footnote-ref-75)
76. The principle of synonym substitution, which would be the only possible principle applicable in the first example, can’t however explain the changes quoted, because the expressions used are not synonyms; for the second example there isn’t any principle that could explain it, because it’s clearly a case of enrich of information. [↑](#footnote-ref-76)
77. Ibid., p. 152. [↑](#footnote-ref-77)
78. The idea of schematic processing could only be admitted in the basic sense of registering what fits in more closely with our personal and cultural point of view; however, this is not the principal meaning of the notion, which is why I claim that it is not compatible with Bernecker’s account overall. [↑](#footnote-ref-78)
79. Malcolm (1977), p. 79. [↑](#footnote-ref-79)
80. Like the same Bernecker says. See p. 153. [↑](#footnote-ref-80)
81. Or the past representation. We’re not going to enter in this discussion which, even if it is important, it is not well developed in the authors we treated. [↑](#footnote-ref-81)
82. Lackey (2005). [↑](#footnote-ref-82)
83. Borges (1944). [↑](#footnote-ref-83)
84. Leaving aside the case of memories of events or episodic memories, which are more obviously incompatible with these reproductive memory conception. [↑](#footnote-ref-84)
85. Matthen (2010). [↑](#footnote-ref-85)
86. We add « images » because in Matthen account, as well as in other authors, such as Fernandez (2008), episodic memory is considered as imagistic whereas semantic memory can be expressed in sentences or in images that are not self-involving. For Matthen, unlike Burge, imagistic content is one form of propositional content, along with sentences. Nevertheless, it would be difficult to see from the examples of imagistic semantic memory given by Matthen in what sense they could be propositional: the look of the Mona Lisa, the sound of a minor scale on a piano, the opening bars of Beethoven’s Fifth Symphony. But I will not go into the discussion of the correctness of the typology proposed. [↑](#footnote-ref-86)
87. Ibid., p. 7. [↑](#footnote-ref-87)
88. Ibid., p. 7. [↑](#footnote-ref-88)
89. Ibid., p. 9. [↑](#footnote-ref-89)
90. Michaelian (2010), p. 3. [↑](#footnote-ref-90)
91. Ibid. [↑](#footnote-ref-91)
92. Ibid., p. 2. [↑](#footnote-ref-92)
93. Ibid., p. 13. [↑](#footnote-ref-93)
94. An idea whose inspiration clearly comes from Lackey’s article (2005). However, I disagree with this characterization of memory as generative of the epistemic status of the belief as well as of the belief itself. Michaelian nonetheless uses the idea of generation in a different sense, because what is generated is not only “beliefs” but “content” in a broader sense, which besides can be integrated not only to the final representation but also to the trace. [↑](#footnote-ref-94)
95. Unlike what Michaelian believes, Matthen’s account seems to be compatible only with the addition of content at retrieval (content that would be external to the trace), but not with the possibility of modification or transformation of the content of the trace at the moment of reconsolidation. [↑](#footnote-ref-95)
96. As I have already noted, this topic is not treated here due to its complexity. [↑](#footnote-ref-96)
97. Martin & Deutscher would be an exception, because as I noted above, they consider that memories can be represented in various ways; however, their notion of “representation” is not defined and is used in a extremely vague way: as Malcolm points out: “'Representation' is a more satisfactory term than 'image' for a memory theorist, since its greater vagueness gives his thesis better protection against refutation.” (1977) p. 67. [↑](#footnote-ref-97)
98. Malcolm (1970), pp. 65 and 67. [↑](#footnote-ref-98)
99. See Bergson (1896) ; Kierkegaard (1843). [↑](#footnote-ref-99)
100. Sutton (2009), p. 229. [↑](#footnote-ref-100)
101. Only in Lackey’s article is it explicitly suggested that at the moment of encoding events and facts could be the object of a different kind of representation that goes beyond the propositional or imagistic one and that would not necessarily be conscious. [↑](#footnote-ref-101)
102. Malcolm (1977), chapters 5 and 9. [↑](#footnote-ref-102)
103. Ibid., p. 244. [↑](#footnote-ref-103)
104. As I have already noted, when it comes to Martin and Deutscher’s account, it’s even more difficult to understand this isomorphism or similarity if we add the requirement of an isomorphism or similarity between the structure or content of the trace and of the mental representations and those of the facts or events of the external world. [↑](#footnote-ref-104)
105. Malcolm (1977), pp. 163-164. [↑](#footnote-ref-105)
106. For a very complete and detailed list of memory analogies : Roediger (1980), p. 233. [↑](#footnote-ref-106)
107. Neisser (1988), p. 364. [↑](#footnote-ref-107)
108. Roediger (1980), p. 232. [↑](#footnote-ref-108)
109. Ibid., p. 237. [↑](#footnote-ref-109)
110. See Collins & Hay (1994), p. 205. [↑](#footnote-ref-110)
111. Koriat. & Goldsmith (1996), p. 6. [↑](#footnote-ref-111)
112. See Anderson (1983b), p. 87. [↑](#footnote-ref-112)
113. Quillian (1962), p. 17. [↑](#footnote-ref-113)
114. Quillian (1968) . [↑](#footnote-ref-114)
115. The aim of considering that the nature of the nodes are “properties” and not directly words, even if his analysis is focused on word meaning, is to leave open the possibility of a single store of information underlying semantic memory and spatio-visual memory, their difference being not in the structure of the information store, but rather in the way that the static information of that store is used. [↑](#footnote-ref-115)
116. Collins & Quillian (1969), p. 241. [↑](#footnote-ref-116)
117. See for example Rosch (1975). [↑](#footnote-ref-117)
118. See Rips, Shoben & Smith (1973). For an extended and reformulated theory of Quillian’s model see for example Collins & Loftus (1975). [↑](#footnote-ref-118)
119. Raaijmakers & Schiffrin (1981). [↑](#footnote-ref-119)
120. Ibid., p. 94. [↑](#footnote-ref-120)
121. Anderson (1983b), pp. 5-6. [↑](#footnote-ref-121)
122. Ibid., p. 19. [↑](#footnote-ref-122)
123. In fact, Lebiere & Anderson developed in an article of 1993 a connectionist implementation of the ACT-R theory. [↑](#footnote-ref-123)
124. Anderson (1983), p. 71. [↑](#footnote-ref-124)
125. Ibid., p. 177. [↑](#footnote-ref-125)
126. Ibid., p. 51. [↑](#footnote-ref-126)
127. Ibid., p. 58. [↑](#footnote-ref-127)
128. Ibid., p. 72. [↑](#footnote-ref-128)
129. Anderson (1983a), p. 13. [↑](#footnote-ref-129)
130. Anderson (1983b), p. 78. [↑](#footnote-ref-130)
131. Ibid., p. 79. [↑](#footnote-ref-131)
132. Idem. [↑](#footnote-ref-132)
133. Ibid., p. 91. [↑](#footnote-ref-133)
134. Ibid., p. 198. [↑](#footnote-ref-134)
135. Ibid., p. 201. [↑](#footnote-ref-135)
136. Ibid., p. 191. [↑](#footnote-ref-136)
137. Ibid., p. 193. [↑](#footnote-ref-137)
138. We have to remember that, given the examples that Anderson provides, elaborations, including schemas, don’t have to present a semantic abstract format, because they could also be simple past episodes that just have a certain analogy with the actual episode that is encoded or remembered. [↑](#footnote-ref-138)
139. McClelland (1981). [↑](#footnote-ref-139)
140. Ibid., p. 170. [↑](#footnote-ref-140)
141. Ibid., p. 171. [↑](#footnote-ref-141)
142. Except if the generalization is a fact learned from a book, internet, what someone told us, etc. Anyway, this doesn’t change the point we wanted to underlie. [↑](#footnote-ref-142)
143. McClelland & Rumelhart (1985). [↑](#footnote-ref-143)
144. Ibid., p. 161. [↑](#footnote-ref-144)
145. McClelland and Rumelhart (1985), p. 160. [↑](#footnote-ref-145)
146. Ibid., p. 161. [↑](#footnote-ref-146)
147. Ibid., p. 162. [↑](#footnote-ref-147)
148. Idem.. [↑](#footnote-ref-148)
149. Ibid., p. 164. [↑](#footnote-ref-149)
150. Ibid., p. 165. [↑](#footnote-ref-150)
151. Idem. [↑](#footnote-ref-151)
152. Idem. [↑](#footnote-ref-152)
153. Ibid., p. 172. [↑](#footnote-ref-153)
154. Idem. [↑](#footnote-ref-154)
155. Ibid., p. 174. [↑](#footnote-ref-155)
156. See Jacoby (1983a, 1983b). [↑](#footnote-ref-156)
157. See, for example, Morton (1979). [↑](#footnote-ref-157)
158. See Collins & Hay (1994). [↑](#footnote-ref-158)
159. Elman (1990), who proposes a model that make use of context units (Jordan 1986) and tries to develop representations of time that are distributed and implicit in the network dynamics. [↑](#footnote-ref-159)
160. I refer to “memory” in general; I do not intend to make an allusion to the specific concept of multiple “memory systems”. [↑](#footnote-ref-160)
161. Collins & Hay (1994), p. 206. [↑](#footnote-ref-161)
162. See Silva (2007), Sara (2007) and Ledoux (2007). [↑](#footnote-ref-162)
163. See for example, Nadel (2007) & Nadel and Moscovitch (1997). [↑](#footnote-ref-163)
164. For Spear (2007), nonetheless, it’s unlikely that all the forms of memory require reconsolidation after retrieval. [↑](#footnote-ref-164)
165. Sara (2007), p. 188. [↑](#footnote-ref-165)
166. Gardiner (2007). [↑](#footnote-ref-166)
167. Nadel (2007), p. 180. [↑](#footnote-ref-167)
168. Tulving & Thomson (1973). [↑](#footnote-ref-168)
169. LeDoux (2007). [↑](#footnote-ref-169)
170. McClelland and Rumelhart (1983), pp. 183-184. [↑](#footnote-ref-170)
171. Smith (2007). [↑](#footnote-ref-171)
172. See, for example, Tulving (1997). [↑](#footnote-ref-172)
173. For a list of examples of connectionist models see Collins & Hay (1994), pp. 212-214. [↑](#footnote-ref-173)
174. Tulving (2007), p. 67. [↑](#footnote-ref-174)
175. Moscovitch (2007), p. 21. See also Schacter (2007). [↑](#footnote-ref-175)
176. Fodor and Pylyshyn (1988). [↑](#footnote-ref-176)
177. See Bechtel (1983). [↑](#footnote-ref-177)
178. Sutton (2009), p. 229. [↑](#footnote-ref-178)
179. See Campbell (2006). [↑](#footnote-ref-179)