RECONSTRUCTING MEMORY

Felipe De Brigard

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill
2011

Approved by:

Jesse J. Prinz

William G. Lycan

Dorit Bar-On

Kelly S. Giovanello

Daniel C. Dennett

ABSTRACT

FELIPE DE BRIGARD: Reconstructing Memory

(Under the direction of Jesse J. Prinz)

According to the received view, memory is a cognitive system the function of which is to store, preserve, and accurately retrieve personal-level representations of past experiences. Philosophers who hold this view typically explain cases of false memories either as mental events that are not produced by memory or as the product of a memory system that is malfunctioning. However, research in the cognitive psychology and neuroscience of false memory presents a challenge to this view, as it strongly suggests that memory distortion is not only a common and pervasive phenomenon but also the result of a well-functioning memory. In my dissertation I argue that in order to make sense of this evidence, we need to reject the received view. In particular, I argue that remembering isn't the retrieval of personal-level perceptual representations, but rather the reconstruction of incomplete sub-personal level sensory representations. Also, I argue that the content of our memories is not carried by a single representation that is preserved through time—a memory trace. I offer instead an account of memory traces according to which they are dispositional properties of neural networks to recreate the mental situation one was in during the original perception.

In fleshing out what this particular disposition actually is, remembering is then explained not as the retrieval of a stored mental representation but rather as the act of reconstructing the mental situation one was in during the original perception by filling-in incomplete sub-personal memory traces. This analysis leads me to argue in favor of a view of remembering in which being aware of the content of a memory consists in covertly attending to a reactivated sensory representation. The final claim I argue for has to do with memory's function: I suggest that memory is not for the reproduction of stored mental representations. Instead, remembering is a sub-operation of a larger cognitive system the function of which is to produce probable episodic counterfactual thoughts—thoughts of what could have likely happened in the past—in the service of guiding and regulating our thoughts about what may happen in the future.

To my first love, Adrianne, and to my new love, David.

TABLE OF CONTENTS

1. Introduction

One hundred years ago, psychology and philosophy professor G.F. Stout defined *memory* in his influential "A Manual of Psychology" as the revival of ideas, insofar as it is "merely reproductive, and does not involve transformation of what is revived in accordance with present conditions". Furthermore, he claimed that such revival "requires the objects of past experiences to be re-instated as far as possible in the order and manner of their original occurrence" (Stout, 1915: 575). After a century of scientific research we now know that such an account is most likely false or, at best, imprecise. Anthropological, psychological and neurological evidence strongly suggests that memory is not a reproductive but a highly reconstructive process, which is critically dependent upon current conditions of recall, and susceptible to several kinds of distortions, many of which do not respect the order or the manner of their original occurrence. Surprisingly, philosophy has been largely insensitive to this evidence. Given the pivotal role that memory plays in our mental life, and given the recurrence with which memory features in many philosophical theories, this is a serious oversight. Reconsidering memory under the light of such evidence has significant—even surprising—repercussions for philosophy. Exploring some of these repercussions is the main purpose of my dissertation.

According to the traditional philosophical view—very much in line with Stout's—memory is regarded, first, as a cognitively isolated faculty: it can either send or receive

information from low-level (e.g., perception) and high-level systems (e.g., reasoning). As such, no part of the system of memory is shared by any other system; when memory gets damaged all other systems remain unaffected. Second, many supporters of this view assume that remembering is a propositional attitude. From this perspective, remembering is understood in terms of a subject relating to a proposition—whether intra- or extra-mental—with the right memorial attitude. Finally, the traditional view holds that memory is primarily reproductive, and that its function is to recapitulate the past as faithfully as possible. Therefore, according to the traditional view, to falsely remember something—as when we mistake a memory for an imagination—should be regarded as a case of memory's malfunction. In contrast, I defend a view according to which memory is not an isolated, but an integrated system; it shares some of its essential components with other cognitive systems, predominantly perception, attention, and imagination. Second, I claim that memory isn't reproductive. While a memory experience phenomenologically presents to the subject as the re-instatement of a previous event, in reality there is nothing that is literally brought back to mind. Third, I argue that remembering shouldn't be understood as a relation between a subject and a proposition, and that the content of our (episodic) memories aren't propositional. Finally, I argue that instead of thinking of cases of false memories as failures of memory, the pervasiveness of this phenomenon is better explained by a theory according to which memory's function is not that of faithfully remembering the past but rather of flexibly using the perceptual components of past experiences to build personal counterfactuals (i.e. thoughts of what could have happened in my past) in the service of episodic foresight (i.e. images of what may happen to me in the future).

The philosophy of memory is sharply divided when it comes to answering the question about *what* we remember when we remember. Two answers have been traditionally suggested. The first of these views is known as *representationalism*. It claims that what you remember is a mental representation whose content is the past event. When you remember your first kiss, for instance, what you are aware of is not the event of kissing someone *per se*, but rather a mental representation depicting such an event. This retrieved mental representation is identical to—or it is caused by—a memory trace left by the original experience. In contrast to this view is *direct* or *naive realism*, which suggests that what you are aware of when you remember your first kiss is not a mental representation of any sort, but rather the event itself—just as you are aware of an object itself, not a mental representation, when you perceive it directly. Direct realists have put forth a number of arguments against the claim that remembering involves memory traces. In the second chapter I tackle one of these prominent arguments, and I offer reasons to believe that memory traces are real and that they play a necessary role in the explanation of recollection.

However, unlike traditional representationalists, I don't believe that a memory trace is tantamount to the personal-level representation one is aware of when remembering. Memory traces should not be understood as mental representations that, having been entertained during perception, are put aside in some sort of unconscious storehouse waiting to be recovered during recollection. Instead, I claim that memory traces are to be understood as multi-level neural mechanisms that acquire the dispositional property to reinstate, more or less, the activation state they were in during encoding at the time of retrieval. The neural activity that gave rise to the encoded

experience is thus reconstructed during retrieval when the appropriate cue triggers the right memory trace. As a result, the view I offer in the next chapter blurs the distinction between the memory trace left by a particular experience, and the processes of encoding and retrieving the memorial content representing that particular experience.

Thinking of memory traces in this way invites us to reconsider the nature of the intentional content of our memories. As I mentioned, the traditional view considers remembering as a propositional attitude. Accordingly, to remember is customarily understood as a relation between a rememberer and a proposition conveying the content of the memory. In the third chapter I offer two arguments against this perspective. On the one hand, I claim that a relational account of remembering encounters certain problems when it comes to explaining the way in which propositions figure in our psychological explanations. On the other hand, I argue that the content of our memories may not be propositional. By this I don't mean that our memories fail to represent the events they are about; they certainly do. However, they do so in a way that suggests degrees of accuracy, rather than simply true or false—as it is assumed by partisans of the propositional attitude account.

By rejecting the view that the content of our memories is propositional, I leave open at least two distinct and important questions. The first question has to do with joint reminiscing. According to the propositionalist view, to be in the same mental state as another person is for both of them to entertain the same proposition with the same mental attitude. Without propositions, then, how can we explain the pervasive phenomenon of joint reminiscing? In the fourth chapter I offer an account of joint reminiscing that disposes of propositional contents. The analysis I put forth seeks to explain not only the

way in which our brain allows us to retrieve the content of our memories, but also the mechanisms by means of which those contents become conscious. Such mechanisms, I argue, are essentially attentional mechanisms, and I claim that remembering is tantamount to covertly attending to a memorial content. Accordingly, in order for two or more people to jointly reminisce about the same memory, they all must be able to covertly attend to the same memorial content.

Finally, the second question left open by my rejection of propositional contents from the analysis of remembering concerns the admission of false and distorted memories as genuine cases of remembering. The idea that such is the function of memory goes back at least to Aristotle (Sorabji, 2006). Notwithstanding such eminent origin, I think this idea is partially wrong: if memory were reconstructive and sometimes inaccurate, wouldn't that speak against memory's function of recording and reproducing the past? In the fifth and last chapter, I present a picture of memory as an integral part of a larger system supporting not only thinking of what *was* the case and what potentially *could be* the case, but also what *could have been* the case. As a result, I argue that the function of memory is to permit the flexible recombination of experiences into possible past events that didn't occur, in order to assist in creating mental simulations of possible future events. This picture of memory will necessarily force us to reconsider memory's function. For now remembering the past is, at most, only a secondary concern for memory. The advantage is that reconstructing memory under this new light eliminates the problem of squaring its function with the fact that our memories are sometimes inaccurate.

2. The ontological status of memory traces

Doubt about traces is to be dispelled by reference to the sheer difficulty of psychological explanation without them.

(John Sutton, 1988: 300)

1. Introduction

The notion of memory trace is probably as old as our interest in understanding memory. However, questions as to whether such a notion denotes a real entity, and what its nature may be, are still a matter of debate in the philosophy and psychology of memory. In this chapter I present a model that may help understand the nature of memory traces, and I offer an argument in favor of their existence inspired by what has been called "experimental realism" (Cartwrtight, 1983; Hacking, 1983). Briefly stated, I suggest that memory traces are multi-level neural mechanisms, and I claim that our capacity to intervene with their mechanistic operations in order to affect subsequent recall gives us epistemic warrant to believe in their existence.

To that end, I need to show that the question about the reality of memory traces is a particular instance—or, at least, that it shares the dialectical structure—of the more general debate in the philosophy of science about the reality of the unobservable entities supposedly referred to by our theoretical terms. I do that in the second part of this chapter, where I claim that arguments in favor of the reality of memory traces are usually

inferences to the best explanation (IBE). In particular, I claim that realists about memory traces argue that positing memory traces as causal intermediaries between the experienced event, *Ex*, and the subsequent mental state of remembering the event, *Rx*, offers the best explanation as to how *Ex* causes *Rx*. On the other hand, anti-realists about memory traces suggest that the postulation of memory traces is not required to explain how *Ex* causes *Rx*. They argue that the main motivation for positing intermediate causal traces stems from the realist's rejection of causal explanations involving action at a temporal distance. By way of arguing in favor of causal explanations allowing spatiotemporal gaps between cause and effect, anti-realists about memory traces try to undercut the main motivation for the postulation of memory traces.

In the third part, I argue against this anti-realist counter-argumentative strategy, not by way of showing that action at a distance is not possible, but rather by suggesting that the mere acceptance of action at a distance still does not give us the best possible *causal* explanation for recollection. More specifically, I argue that even if one accepts the possibility of causal explanations involving action at a distance, there are still many causally related questions about recollection for which that sort of explanation is insufficient. I offer instead a model for the causal explanation of recollection using an interventionist framework (Woodward, 2003). I explain how such an account would fare better than the mere action at a distance account—and, incidentally, than the mere causal account—when it comes to many causally relevant questions about recollection. As anticipated, this proposed model requires the existence of causally relevant memory traces.

To substantiate the model, I offer—in the fourth section of this chapter—a mechanistic interpretation of the postulated memory traces drawing from recent data in the neuroscience of learning and memory. I explain what, according to these empirical data, memory traces may be, and how they can behave causally as suggested by the interventionist model. Finally, in the fifth and last section, I explain how neuroscientists have in fact manipulated, and predict how they can manipulate, memory traces in order to produce differential effects in subsequent recollection. This final observation leads me to suggest that our capacity to manipulate memory traces gives us reason to believe in their existence.

2. Memory traces as theoretical posits

Although the notion of memory trace is widely used in philosophy of memory, as well as in cognitive psychology and neuroscience, there is substantial disagreement as to what exactly is meant by it. In its most general form, 'memory trace' is used in reference to events or processes that exist during a period of time, t_2, between a time, t_1, in which a subject, *S*, experiences a particular event *x*, *Ex*, and a subsequent time, t_3, in which *S* remembers or recollects *x*, *Rx*. Additionally, some philosophers of memory tend to think that in order for something to count as a memory trace, three conditions must obtain: (1) the memory trace must play a causal role in the recollection of the event it is a trace of, (2) it must retain the mental content entertained during the remembered event, and (3) it must be structurally similar or isomorphic to what is remembered (e.g., Malcolm, 1977; Bernecker, 2008).

Of these conditions, only the first one is relatively uncontroversial. With the advent of connectionist models of cognition (e.g., Rumelhart et al., 1986) and the discovery that neural networks are constantly redeployed for different cognitive tasks (e.g., Anderson, 2010), cognitive scientists have questioned the claim that there is a meaningful way of talking about memory traces *preserving* any sort of mental content—as opposed to, say, *reconstructing* it during retrieval (McClelland, 2010). Additionally, both philosophers of memory and cognitive scientists doubt that there are either conceptual or empirical reasons to believe that even if it makes sense to say that memory traces preserve mental contents, they need to do so in a format that, in any interesting way, is structurally isomorphic to the remembered experience or event (e.g., Rosen, 1975). Notwithstanding the importance of these issues, I want to focus on the more fundamental question about the very existence of memory traces—a question that, as I plan to show, is intimately related to our understanding of condition (1). However, before I get there, I want to explore where the question about the existence of memory traces comes from.

2.1. Two ways of thinking about the existence of memory traces

The notion of memory trace predates the distinction between philosophy and psychology; indeed, it predates the distinction between philosophy and *science*. One of the first references to memory traces comes from Plato's *Theaetetus*, where experiences leaving traces in our memory are compared to seal rings leaving impressions in a wax table. These impressions—the analogy tells us—are representatives of the seal ring, just as memory traces are representatives of the experiences that created them (Plato,

*Theatetus*194 c-e). Philosophical tradition also tells us that Zeno the Stoic and Aristotle embraced the view that perception leaves traces, and that such traces give rise to the memories we later on recover during recollection (Sorabji, 2006; Gomulicki, 1953). Likewise, the appeal to memory traces persists in the views on memory of early modern philosophers, like Descartes and Hobbes, as well as in the views of empiricist philosophers, like Locke, Hume and Mill. Aside from some notorious skeptics (e.g., Reid, 1785/1849), the use of memory traces to explain the phenomenon of recollection was so widespread by the end of the 18th century and the beginning of the 19th century, that it became the received view by the time psychology established itself as an independent discipline.

By the end of the 19th century both philosophers and psychologists seemed to agree that, given the current status of neuroscience, memory traces were merely hypothetical (Russell, 1921). However, philosophers and psychologists disagreed as to *how* to interpret the precise nature of this hypothesis, and the way one should go about verifying it. As a result, the quest for the ontological status of memory traces became the object of two different and relatively independent inquiries. On the one hand, philosophers saw the postulation of memory traces as a *theory-independent hypothesis*. Memory traces were hypothetical precisely because their acceptance within a theory of memory was at stake. As a result, they thought that the first step in order to know whether or not there are memory traces was conceptual: one needed to find out whether the notion of memory trace was at all required for our correct understanding of memory. On the other hand, psychologist thought of memory traces as a *theory-dependent hypothesis*; indeed, they thought of it as a psychophysical hypothesis (James, 1890: 655).

From the psychologist's point of view, memory traces were hypothetical, not because we were questioning whether or not they were required at all for a correct account of memory, but rather because we just didn't know what sort of physical—i.e., neural—entity they could be. As such, psychologists thought that the task of verifying the nature of memory traces was an empirical one; it had to do with finding out the nature and precise location of memory traces in the brain, not with whether or not we were justified in postulating them. Thus, while philosophers like Russell (1921) and Broad (1925) were interested in finding out whether or not we required the notion of memory trace in order to have a full-fledged analysis of our concept of memory, psychologists like Semon (1904/1921)—who coined the term "engram" to refer to memory traces—and James (1890) were in the business of devising theories about the biological and physiological nature of memory traces.

This is not to say that psychologists accepted the existence of memory traces by fiat. On the contrary, with the advent of methodological behaviorism, the notion of memory trace fell in disrepute (Watson 1930, Skinner 1953). Any mention of 'memory traces'—indeed, any mention of 'memory' as opposed to 'learning'—was practically jettisoned from psychological writings, and those who kept searching for the engram reached rather pessimistic conclusions. In 1950, Karl Lashley—who was trained as a behaviorist by J.B. Watson—published his famous paper *In Search of the Engram*, in which he declared that "it is not possible to demonstrate the isolated localization of a memory trace anywhere within the nervous system" (Lashley, 1950). Many interpreted

the results therein presented, as well as Lashley's view, as substantial evidence against the existence of memory traces.[1]

In the mid-1960s, however, two relatively contemporaneous discoveries resurrected psychologists' hopes for finding memory traces. The first one occurred in 1957, when Brenda Milner described the case of H.M., an individual who had sustained a bilateral resection of the medial temporal lobes, four years earlier, as a result of an intractable epilepsy (Scoville & Milner, 1957). The surgery left H.M. completely unable to store new information, and it impaired his recollection of recent memories, while apparently sparing all other intellectual abilities. This observation strikingly opposed Lashley's views, for it showed that there was a clear dissociation between brain areas that were required for the creation of new memories, and the retention of recent ones, and brain areas that were not. This important observation confirmed the fact that without a hippocampus you simply cannot remember new experiences.

The second discovery took longer, but it was equally influential. Work in synaptic plasticity in hippocampal cells led neurophysiologist to insert a tetanus in the pre-synaptic membrane in order to increase the speed and the repetition of electrical stimulation, which in turn allowed them to extend the life of the electrical signal they were recording in the post-synaptic cell. Researchers in Per Andersen's neurophysiology lab in Oslo began to observe what appeared to be a correlation between the frequency and duration of the tetanus' burst in the pre-synaptic cell and the length and enhancement of the post-synaptic response, ranging from a few seconds (Green & Adey, 1956) up to about ten minutes (Andersen, Bruland & Kaada, 1961). This experimental trick soon

[1] I believe that a careful examination of Lashley's work reveal that this interpretation is way out of proportion. I am just following the traditional story here.

became an object of research in and of itself, as they began to wonder about the underlying mechanisms that allowed hippocampal cells to retain their synaptic potentiation long after the electric stimulus was removed. The first description of the underlying mechanism of this phenomenon, known now as long-term potentiation (LTP), was offered by Tim Bliss and Terje Lomo in 1973 (see Craver, 2003). The discovery of a neural mechanism that could preserve the effects of a stimulus once removed, and the fact that such mechanism happened to be found in a region that was demonstratedly necessary for the formation of new memories, gave a new life to the research on memory traces. Today, most neuroscientists accept their existence, and the quest for the engram remains an active line of research (e.g., Thomson, 2005).

2.2. From causation to memory traces

For the philosopher, however, these results do not count as evidence until the philosophical question about the justification for the postulation of memory traces gets an affirmative answer. For, if it doesn't, there is no reason to believe that neuroscientists are justified in taking memory traces to be the sort of entity that can be empirically discoverable. And, as it happened, during most of the first half of the 20th century, many took the answer to the philosophical question to be in the negative. Russell (1921) and Wittgenstein's (1953) influential views led philosophers to think that our concept of memory did not require a reference to the cause of that which is remembered, much less to a causal *link* between the past experience and its subsequent recollection. They thought that the inclusion of a causal link between the experience and its subsequent recollection was neither a necessary nor a sufficient condition for a successful account of

remembering. This view, initially championed by Ryle (1949) and Benjamin (1956), found its clearest expression in Malcolm's *Knowledge and Certainty* (1963), where he emphatically declared that "our use of the language of memory" (p. 237) carries no implication about the causes of our remembering or about the causal mechanisms involved in our recollections (see also Munsat, 1966). As a result, with no conceptual reasons to accept the postulation of memory traces for a correct account of memory, the idea that memory traces may in fact refer to something real became a thing of the past. To philosophers, neuroscientists were, at best, pursuing will-of-the-wisps.

However, that same year, Martin and Deutscher's celebrated paper *Remembering* (1963) put realism about memory traces back on the table. The main purpose of that paper was to argue against the claim that a causal condition is not required for a proper analysis of our concept of remembering. Their argument is based on cases—some real, some imaginary—of people that had a particular experience x at a certain time, t_1. Then, during some subsequent and arbitrary interval of time, t_2, they forget the event. However, at a later time, t_3, these people do something "for which the only reasonable explanation" (p. 176) is that they experienced x at t_1. The observation that a causal claim is required in order to make sense of their behavior at t_3 as a consequence of the experience at t_1 motivates Martin and Deutscher to claim that "if a person's account of what he saw is not due even in part to his seeing it, it cannot be said that he remembers what he saw" (Martin & Deutscher, 1963: 175-176). Since the examples Martin and Deutscher discuss intuitively fall under our concept of remembering, then they conclude that a person's recollection of x must be due to her having experienced x. Thus, they formulate their causal condition for remembering:

(CC) *S*'s experience of a particular event *x*, *Ex*, causes—or, at least, is causally relevant to—*S*'s subsequent recollection of the event, *Rx*.

There are two points I would like to extract from Martin and Deutscher's analysis. First, their paper brought causation back into the analysis of memory by way of pointing out the explanatory indispensability of the causal condition. Unfortunately, this point isn't stressed enough, partly because Martin and Deutscher's own analysis suggests that CC is simply the conclusion of the following modus tollens:

Argument 1:

(P1) If S's *Rx* is not caused by *Ex*, then S is not *Rx*-ing (Assumption)

(P2) But S is *Rx*-ing (as evidenced by their thought-experiments)

(C) Therefore, S's *Rx* is caused by *Ex*.

But notice that this interpretation may render their argument vacuous. As stated, P1 implies its counterpositive:

*(P1) If S is *Rx*-ing, then S is caused by *Ex*

which is precisely the conditional they want to prove. But, of course, Martin and Deustcher should not want that, for if P1 is supposed to be an *assumption*, then the argument shouldn't prove what they assumed to begin with.

I think a better interpretation is to treat their argument *inductively*, so that the thought-experiments they discuss are to be accommodated by an IBE. Consider one of their well-known thought-experiments (Martin & Deustcher, 1963). A painter is asked to draw an imagined rural landscape. When he's done, his parents recognize the painting as depicting the view from the house they used to live in many years ago. The painter, who has no recollections of the time they lived in such house, claims to have imagined the scene he painted. However, the intuition this thought-experiment is supposed to elicit—and let's assume that it does for the sake of argument—is that the painter is actually painting the scene from memory and not from imagination. Notice that Martin and Deuschter's claim is that we take the painter's envisioning the scene while painting it as a case of remembering (as opposed to imagination) *because* the "only reasonable explanation" for his mentally entertaining that precise scene at t_3 is his having experienced it at t_1, even if he did not remember it at all during t_2. Thus, if we take their argument to be an IBE, then CC would enter as the hypothesis that best fits the data provided by the thought-experiment. Schematically (Lipton, 1990):

Argument 2:

(P1) The painter's case is clearly an instance of recollection.

(P2) The hypothesis CC, if true, would explain why the painter's case is an instance of recollection.

(P3) No other hypotheses can explain why the painter's case is an instance of recollection as well as CC does.

(CC-C) Therefore, CC is (probably) true.

If this is the case, then the appeal to CC is the result of an IBE, as opposed to the conclusion of a deductively valid argument.

The second point I'd like to extract from Martin and Deutscher's analysis is that the existence of memory traces is supposed to follow, as a matter of course, from the acceptance of CC (Martin & Deutscher, 1963: 189). Their argument, which is not terribly straightforward in their paper, can be reconstructed as follows (see Malcolm 1977):

Argument 3:

(P1) *S*'s *Ex* is diachronically separated from S's *Rx*.

(P2) A cause cannot be diachronically separated from its effect (i.e., there is no causation at a temporal distance).

(MTC) Therefore, there *must* be an intermediary causal connection, *Mx*, between *Ex* and *Rx* such that *Ex* is the proximal cause of *Mx* and *Mx* is the proximal cause of *Rx*.

So if *Mx* stands for 'memory trace of x', then the memory trace clause (MTC) tells us that there are memory traces (see Rosen, 1975, for a similar conclusion). But notice, once again, that the argument hinges on an IBE. The idea is that the postulation of memory traces as causal intermediaries between *Ex* and *Rx* allows us to preserve CC without having to accept the metaphysically uncomfortable claim that there is causation at a temporal distance. In other words: memory traces become theoretical posits postulated to

help explain the causal connection between *Ex* and *Rx* without having to accept action at a temporal distance. Thus, Malcolm tells us:

> Presumably the reader will know that memory traces are not entities, states, or processes that neural surgeons have discovered in the course of their investigations of the brain, as dentists discover cavities. The memory trace is what may be called "a theoretical construct". It is something that is inferred to exist from the presence of things that unquestionably exist, such as learned skills, habits, and occurrences of recognition and remembering (Malcolm,1977: 171).

Similarly, John Heil writes:

> "It is important to see that the existence of traces is not supported by independent psychological or physiological evidence. Traces are postulated just because it is thought that their postulation provides an explanation for the phenomenon of memory—and perhaps other psychological processes as well. [Memory traces] are what once were called theoretical entities, devices introduced in the context of a theory to explain some more accessible phenomenon". (Heil, 1978: 62).

It is worth remarking that, at least for Malcolm, memory traces are postulated in virtue of our "abhorrence of 'action at a distance'—in this case, action at a temporal distance" (Malcolm, 1977: 174). Memory traces are conceived as playing the role of "bringing about a memory response: [for] without the existence of a trace there would be a gap in a causal chain and causal action would occur at *a temporal distance*" (Malcolm, 1977: 1979. Emphasis in the original). In sum, the motivation behind the postulation of a memory trace as a theoretical posit is the fact that it constitutes a better explanation of how *Ex* causes *Rx* than the alternative action-at-a-temporal distance account in which *Ex* directly causes *Rx*.

The problem with this conclusion, however, is that it leaves open the following possibility: when it comes to explaining how *Rx* came about as a result of *Ex,* an explanation involving action at a temporal distance could be *at least as good* as an explanation involving memory traces. This is precisely Norman Malcolm's important move in his *Memory and Mind* (Malcolm, 1977). His argument, which is reminiscent of Russell's (1921) defense of "mnemic causation", is that the kind of explanation we usually invoke when talking about remembering does not imply, in any way, that there should be a process mediating *Ex* and *Rx*. Suppose—to use one of his examples—that you tell someone that you saw a boat capsize last week. Now imagine that, for whatever reason, your interlocutor is in disbelief: 'How do you know that?', she asks, to which you reply 'I know because I saw it happen' (Malcolm, 1977; 183). The thought here is that in explaining how it is that you remember the boat capsizing, you are applying a causal claim, just as Martin and Deutscher argued, but it makes no reference to any sort of causal process or state mediating the event perceived and your recollection of it. As a matter of fact, it makes no sense to ask whether you are certain that there was an ongoing causal process between your witnessing the boat's capsizing and your relating the story.

> The strange, irrelevant character of this question shows that there is a familiar use of causal language consisting of such ordinary locutions as "because of", "due to", "the cause of", and the more technical "necessary causal condition", which carries no implication of a causal process filling up the temporal space between the occurrence of a cause of x and the occurrence of x. We can agree with Martin and Deutscher that the language of memory does, in a sense, require a "causal interpretation", but not agree that memory as a causal concept entails the concept of causal process […] Eliminate the assumption of a causal *process*, and *the causal argument* for a memory trace collapses. (1977: 185).

In sum, Malcolm argues that one can accept CC without having to admit the need for postulating memory traces. Causal explanations involving action at a temporal distance are—according to him—perfectly reasonable explanations for psychological phenomena such as remembering, and nothing whatsoever about intermediate causal processes is implied by our use of the concept of remembering. While CC may be an IBE as to how *Ex* causes *Rx*, one does not need to accept the second IBE in which memory traces are postulated; explanations involving action at a temporal distance are as good as those involving memory traces. In the next section, however, I argue that they are not.

3. Intervening memory

In the previous section I claimed that the appeal to memory traces stemmed from the realization that their postulation was required to come up with the best possible explanation as to how *Ex* causes *Rx*. After all, the assumption of causally mediating memory traces avoided the uncomfortable metaphysical pitfalls of causation at a temporal distance. Malcolm's anti-realist reply, however, was that one could accept the claim that *Ex* causes *Rx* without having to be committed to causally mediating memory traces, simply because explanations involving action at a temporal distance are equally good explanations for remembering. As a result, the postulation of memory traces as theoretical entities for an adequate account of remembering was thought to be unnecessary, and the idea of trying to find them empirically was deemed unwarranted. With this move, Malcolm made anti-realism about memory traces, once again, an attractive theory in the philosophy of memory.[2]

[2] Coincidentally, the idea that memory traces may not be required for a successful explanation of how *Rx* can be brought about by *Ex* also received some attention in psychology, as it constituted the backbone of

Notice, though, that Malcolm is not arguing in favor of the possibility of action at a temporal distance as a *metaphysical* claim. Whether or not a cause can bring about an effect after a temporal gap is irrelevant to Malcolm's argument. His point, just as Martin and Deutscher's, is about causal *explanation*. After all, he accepts the IBE motivated by 'Argument 2'. What he rejects is the IBE motivated by 'Argument 3'. And he rejects it, not because he denies P2 as a metaphysical claim, but rather because he denies that the acceptance or rejection of causation at a temporal distance has anything to do with successful causal explanations for the phenomenon of recollection.[3] In other words, he does not think that the postulation of intermediary causal processes adds anything to our account of how *Rx* was brought about as a result of *Ex*. This, however, is what I think Malcolm gets wrong, for I am not sure how explanations involving action at a temporal distance can really satisfy our explanatory necessities when it comes to various causally relevant questions about memory and remembering. If we *only* care about experiences causing successful recollections, as Martin, Deutscher and Malcolm do, *maybe* a case can be made to the effect that action at a temporal distance is all we need to accept in order to furnish satisfactory causal psychological explanations. But successful recollection is *not* the only thing we care about when we demand causal explanations for our memories. We often want to know, for instance, why is it that a person, having experienced an event, can nonetheless *fail* to remember it. Additionally, we may want to know why, given that a subject experienced a particular event, she only managed to remember *part* of it, or why

---

the ecological approach to remembering (Gibson, 1979; Turvey & Shaw, 1979; Michaels & Carello, 1981). Much of what I say here could easily apply to this view. For a nice criticism of the ecological approach to memory, which I find very congenial to the spirit of this paper, see Sutton, 1998, part IV.

[3] To put it a la van Fraassen (1980): for Malcolm, an explanation involving action at a temporal distance is all one needs to save the phenomenon of recollection, so there is no reason to believe in the reality of the intermediary unobservable events supposedly referred by our notion of memory trace.

she remembered it *distortedly*. Moreover, sometimes we wonder whether it is possible to *facilitate* or to *hamper* our subsequent recollection of an event after having witnessed it. To put it succinctly, we often wonder whether it is possible to intervene in the alleged causal connection between *Ex* and *Rx*.

Let me offer an analogy to drive my point home. Consider a case in which someone consumes cyanide at t_1 and then dies at t_3. A natural way of describing the event is to say that the person died as a result of her consuming cyanide; that the ingestion of cyanide caused her death. If all we want to know is why she died at t_3, alluding to her consuming cyanide at t_1 may be a sufficient explanation. The same goes for remembering. As I stressed, Malcolm's examples (as well as Martin and Deutscher's) only pertain to successful recollections of past events. After all, the motivation behind the IBE that lead to the acceptance of CC is simply that we cannot make sense of a particular *successful* memory retrieval behavior at *t3* unless we accept as its cause having the relevant experience at *t1*. A similar IBE is at work when, upon seeing a dead body exhibiting the distinctive signs of cyanide poisoning, a coroner alludes to the person's prior ingestion of cyanide as a causal explanation of his death. In such a case, asking the coroner whether or not he's certain that a causal process was going on between the person's ingestion of the cyanide and his eventual death would seem as awkward as asking whether or not you are certain that a causal process was going on between your witnessing an event and your relating it afterwards. Here, alluding to your having witnessed a boat capsizing—to borrow Malcolm's own example—may be enough of an explanation as to why you remember it, the same way in which alluding to cyanide ingestion may be enough of an explanation for the person's death.

But suppose that, right next to the dead body, there is another person who also ingested cyanide but *failed* to die. Let's assume that she exhibited some of the symptoms—shortness of breath and pink skin color—but none of the lethal ones, like pulmonary edema and cardiac arrest. Again, in this case, cyanide ingestion can explain the person's symptoms. For example, if someone asks why her skin is pink, one can rightly say that it is due to her having ingested cyanide. But then an obvious question arises: given that both subjects ingested cyanide, why is it that only one of them died while the other *failed* to die? Now, I take it, talk of intermediary causal processes becomes necessary. The only way in which one can explain why, given the same initial conditions, one person died while the other person failed to die, is by way of alluding to some difference in the causal process that occurred between the cyanide ingestion and the subsequent symptomatic behavior. One possibility is that the person who survived had increased levels of hydroxocobalamin in her blood due to, say, excessive consumption of vitamin B_{12}. As a result, the ingested cyanide preferentially bonded molecules of hydroxocobalamin, leaving the hemoglobin's cytochrome oxidase less affected—which would explain why her levels of blood oxygenation were enough to elicit shortness of breath and skin coloring but *not* pulmonary or cardiac arrest. As I discuss below, there are other possible explanations. The point, though, is that when it comes to explaining the differential effects of cyanide ingestion in these two people, any successful causal explanation is going to involve intermediary causal processes.

The same is true in the case of memory. Consider a small modification of Malcolm's example. Suppose that you weren't alone when you witnessed the boat capsizing. You were with your friend Mary. Both you and Mary were side by side when

the event occurred, both of you were looking at the event, and both of you have roughly the same visual acuity. However, only you remember the event later on. Now, when you wonder why is it that you remember the event while Mary fails to remember the same event, even when both of you witnessed it, appealing to an intermediary causal process is the natural way to proceed. One may say, for instance, that Mary wasn't paying attention, or that she has a short-term memory problem, or perhaps that she has seen so many boats capsizing lately that she cannot remember just that one. Of course, one may allude to some more "organic" explanations; one may say, for instance, that Mary was given an amnestic drug right after she witnessed the event, or that she suffers from some kind of neurodegenerative disease, or even that her medial temporal lobes were damaged at some point after having witnessed the boat capsizing. Whatever the story we tell, it is going to involve a reference to intermediate causal processes that differed between her case and yours.

Notice that the point I am making does not hinge on our knowledge of the neural mechanisms by means of which memories get consolidated and further retrieved—that part of the story will come later. My point so far is about the necessity of alluding to intermediate causal processes in order to reach the best causal explanation of an *unsuccessful*—versus a *successful*—case of remembering. In other words, a causal explanation that does not make reference to intermediate causal processes won't be able to account for the differential effects between cases of successful and unsuccessful recollection. This means that, when it comes to accounting for differential effects during recollection, a causal explanation that does not involve intermediate causal processes won't be as good a causal explanation as one involving intermediate causal processes.

Thus, Malcolm is wrong when claiming that, when it comes to recollection, explanations involving action at a temporal distance are explanatorily on par with those that posit intermediate processes.

To be sure, this argument can also be made when the differential effect involves *improved*—as opposed to impaired—recollection. Suppose that Mary did not fail to remember the witnessed event but she actually remembered it better than you did. Unlike you, she remembered—let's say—that there was a red fender on the starboard side when the boat capsized. Again, other things being equal, any successful explanation is going to involve some reference to intermediate causal processes that differed between you and Mary. These processes can be as simple as closely attending to the fender while witnessing the event, or as complex as having received a dose of strychnine right after seeing the boat capsizing[4]. I think the same goes for other differential effects, not only between subjects but also within subjects. For example, someone may adduce lack of sleep when trying to explain why she failed at a particular test that later on, after a good night of sleep, she can pass with no trouble. The fact of the matter is that we often allude to intermediate causal processes when we offer explanations of differential effects in recollection.

Malcolm is wrong, then, in thinking that causal explanations involving action at a temporal distance are explanatorily equivalent to those involving intermediate causal processing. He isn't entirely to blame, though. The root of the problem, I think, lies in interpreting CC as stating that a reference to *Ex* may be a sufficient condition for explaining how *Rx* came about. Martin and Deutscher also share this assumption, for they

[4] Administering certain chemical compounds such as strychnine during memory consolidation has shown to enhance retention in some mammals (e.g., McGaugh & Krivanek, 1970).

appeal to memory traces via the second IBE stated in Argument 2—the second premise of which Malcolm rejects. But this is the wrong way to introduce memory traces. What Martin and Deutscher should have said is that CC *and* memory traces are a package deal. More precisely, what they should have said is that appealing to the past event *Ex* alone does not constitute the best causal explanation for *Rx* (save, perhaps, in the very circumscribed and highly under-described cases of successful recollection that Malcolm discusses). The past event is *part* of the causal explanation[5], but on most occasions, as in the cases of differential effects just reviewed, the appeal to memory traces is also required, not as a fallout of accepting the past event as the cause of *Rx*, but as a resource to explain the psychological effect itself. I suggest, therefore, a causal condition that incorporates memory traces:

(CC + MTC) *S*'s experience of a particular event *x*, *Ex*, plus a memory trace of *x*, *Mx*, cause—or, at least, are causally relevant to—*S*'s subsequent recollection of the event, *Rx*.

Now, how is it that memory traces become causally relevant when it comes to explanations of differential effects in recollection? In order answer this question, I will rely on James Woodward's recent version of a manipulability theory of causation (see

[5] My argument, so far, implies that *Ex* is not sufficient for *Rx*. In chapter 5 I will also claim that, in a certain sense, it isn't necessary either, as there are certain cases of false and distorted memories in which one remembers an event that did not occur, or that did not occur exactly as remembered. However, as I explain in that chapter, there are certain constraints to this claim, as even for false and distorted memories there is still a causal path between *Ex* and *Rx*.

Woodward 1997, 2000, 2002 and 2003)[6]. According to his view, causes are considered devices for manipulating and controlling effects. Causal explanations explain because they convey information about the way in which one could potentially manipulate or control a certain effect by intervening on a previous event we take to be its cause. Thus, successful causal explanations are used to answer what Woodward (2003) calls what-if-things-had-been-different questions: "the explanation must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways" (Woodward, 2003: 11). In the case of cyanide poisoning, for instance, understanding why one person died while the other one survived requires understanding what it is that we could have done in the case of the person that died to affect the result that occurred in the case of the person who did not die. In other words, we want to know whether there was something one could have done between t_1 and t_3 to prevent her death. Might it have been possible that even though she ingested cyanide at t_1 we could have done something at t_2 in order to prevent her death at t_3?

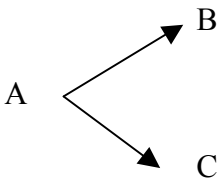
As it turns out, there are ways in which one can prevent death by cyanide poisoning. For instance, we know that cyanide, when dissolved in water, inhibits cytochrome oxidase blocking electron transport, which in turn decreases the amount of oxygen in the blood. This condition causes lactic acidosis, whereby the pH of the hemoglobin is reduced and it starts building up D-lactate, which rapidly damages organic tissue—especially in our lungs and stomach—thus leading to one's death. As a result, a person's death after ingesting cyanide is potentially preventable at several points during

[6] Woodward's view, of course, is not the only view about why casual explanations explain. I will not defend his view against the usual contenders, but the reader is welcome to check Woodward & Hitchcock, 2003a and 2003b, for that purpose.

the process. Most typically, one could administer nitrites to turn hemoglobin into methahemoglobin, which is preferentially bonded by the cyanide. The bonding of cyanide and methahemoglobin creates cyanmethahemoglobin, which in turn can be treated with sodium thiosulfate to convert the cyanmethahemoglobin into hemoglobin, sulfites, and thiocyanate, the last of which can be secreted through urine without further damage to the organism. However, other possible manipulations could be potentially implemented, like the use of hydroxocobalamins to artificially increase the pH level in the hemoglobin while eliminating the cyanide, or by finding a mechanism to inhibit the creation of D-lactate.

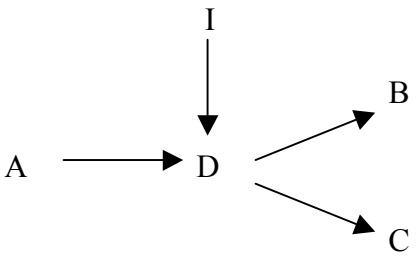
The relevant point is that these interventions—some of which are in fact implemented in medical facilities (like the use of nitrites) and some of which are merely potential (like the use of some chemical agent that could reduce the hemoglobin's pH)—allow us to manipulate the build up of D-lactate. When the levels of D-lactate in the blood reach a certain threshold, a body enters into the physiological condition known as lactic acidosis, which can be lethal. But if one can reduce the levels of D-lactate, lactic acidosis is then prevented and the chances of survival increase. Therefore, according to the manipulation account I am relying on, the immediate cause of death in the cyanide poisoning case just described is the amount of D-lactate in the person's blood. We can tell that because we know that had we intervened to reduce the level of D-lactate in the blood, the person would have merely experienced shortness of breath and skin discoloration.

Figure 1:



Let me put it graphically. As I described the case above (see figure 1), there are two different events, *B* (death via cardiac arrest) and *C* (skin discoloration), which appeared to have been caused by the same event *A* (cyanide consumption). However, as I argued, if we limit our causal explanation to *A*, the differential effect would remain mysterious. So we wonder whether some other event *D* happened between the time t_1 in which *A* occurred, and the time t_3 in which both *B* and *C* occurred, such that it could explain the differential effect. As it turns out, there is: lactic acidosis. We know that *D* is the cause of *B* because we can intervene, *I*, on *D* and prevent the build up of D-lactate, thus manipulating the effect and 'switching', as it were, the causal path from *B* to *C* (figure 2).

Figure 2:



The same, I surmise, occurs with memory. The suggested variation on Malcolm's case unveils a differential effect between a situation in which one remembers the event

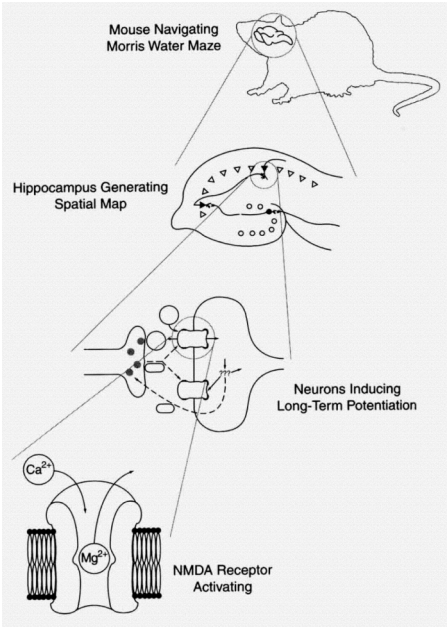
(B) and situation in which one does not remember the event (C), despite the fact that in both situations one has experienced the initial event (A). As in the case of the cyanide poisoning, appealing merely to having witnessed the event does not explain the differential effect. So we wonder whether there is an intermediate event, *D*, such that a proper intervention, *I*, upon it could switch the causal chain from *B* to *C*. In other words, we want to know whether there is an intermediate causal mechanism that could explain why one person remembered or failed to remember a particular event. And we could know that with the appropriate intervention. Enter neuroscience.

4. Memory traces as multi-level neural mechanisms.

In the previous section I argued against Malcolm's anti-realist view, according to which causal explanations for recollection involving action at a temporal distance are at least as good as causal explanations of recollection involving causally mediating memory traces. I claimed that Malcolm's anti-realist accounts are unable to explain differential effects in recollection. Then I suggested an interpretation of the role of memory traces in causal explanations of remembering in terms of Woodward's manipulability theory of causation (2003). Now I would like to suggest a mechanistic model for memory traces that can provide a framework for understanding the way in which certain experimental manipulations conducted by cognitive neuroscientists have actually produced—and could possibly produce—differential effects in recollection. Since this interpretation is largely inspired by Craver's account of multi-level neural mechanisms (Craver 2002, 2007), it is useful to explain what he means by such terms.

According to Craver, a mechanism is a set of entities and activities arranged in specific ways to produce regular changes in a period of time (see Craver 2001; for the original formulation, see Machamer et al., 2000). By 'neural mechanism', therefore, I will refer to the sorts of mechanisms studied in neuroscience. Neural mechanisms typically include entities such as neurons, neurotransmitters, oligodendrocites, hippocampi, brains, etc. They also include activities such as neuronal firing, enzyme release, information processing, brain region activation, etc. The entities and the activities composing neural mechanisms have spatial and temporal organizations that are essential for the mechanism to perform its operations. Finally, the ways in which the mechanism's entities and activities are organized typically compose hierarchies. Each strata of the mechanistic hierarchy is usually called a 'level', so mechanisms that can be decomposed into more than one level are 'multi-level mechanisms' (Craver, 2002; 2007).

Figure 3:



As an illustration of a multi-level neural mechanism, consider Craver's example of a mechanistic decomposition of spatial memory in four levels (figure 3). Each level is the object of study of a relatively independent sub-discipline in the neurosciences, as each level is investigated with a distinctive array of experimental methods. The top level includes entities such as organisms (e.g., mice, humans) and surrounding environments, as well as activities such as discrimination, button pressing and swimming. At this level, experimental psychologists, cognitive ethologists and comparative psychologists study spatial memory using experimental methods like the Morris water maze, radial arm mazes and virtual reality computers. In the second level (one level down) we find entities such as the hippocampus and the enthorrinal cortex, as well as computational activities such as informational transfer and spatial map formation. This level is usually studied by cognitive neuroscientists and neuropsychologists via experimental methods such as event related potentials (ERP), functional magnetic resonance imaging (fMRI), positron emission tomography (PET) scans and several diagnosis assessment methodologies often implemented in clinical settings. The relevance of the hippocampus and the enthorrinal cortex—that is, the entities of the second level—is determined by their dependence on the entities and the activities of the third level. This level includes entities such as granule and pyramidal cells, and activities such as neuronal firing and depolarization. Neurophysiologists and, to some extent, neuroanatomists, study this mechanistic level with experimental methods such as intra- and extra-cellular recording, cell body staining, track tracing and, sometimes, very localized neuropharmacological interventions, like microiontophoresis, whereby the researcher injects small dosages of particular chemical

compounds directly into the neural tissue. Finally, the bottom level consists of molecular mechanisms that include entities such as N-Methyl D-aspartic (NMDA) receptors and Mg^{2+} ions, and activities such as binding and electron releases. Molecular neurobiologists study this level using experimental methods such as pharmacological interventions and gene knockouts (Craver, 2002; 2007)

Although an oversimplification, Craver's spatial memory example highlights an essential feature of any multi-level neural mechanism—including, as I suggest, memory traces. Craver calls it *mutual manipulability,* and it basically specifies a condition of sufficiency for a component to be a part of a multi-level mechanism. According to the mutual manipulability condition, "a part is a component of a mechanism if one can change the behavior of the mechanism as a whole by intervening to change the component *and* one can change the behavior of the component by intervening to change the behavior of the mechanism as a whole" (Craver, 2007: 141). For example, we can tell that LTP in the pyramidal cells in CA1 of a rat's hippocampus is part of the multi-level neural mechanism of the organism's spatial memory because we can intervene its mechanistic operations—by removing NMDA receptors in this location, for instance—thus inhibiting the activity of the place cells, and making it impossible for the hippocampus to form spatial maps. Conversely, we can alter the induction of LTP in CA1 by way of intervening higher neural levels, e.g., severing afferent neural tracts or modifying the rat's behavior.

With the conceptual scaffolding of the mechanistic account, we can go back to our discussion about the existence of memory traces, and ask whether there is a multi-level neural mechanism one can intervene upon in order to bring about differential effects

in recollection. Since most interventionist techniques in neuroscience are relatively new—particularly those that afford controlled manipulations at specific mechanistic levels—the precise nature of such a mechanism is currently poorly understood. However, a number of experimental results are starting to reveal its structure. First, consider manipulations at the molecular level. In a now classic study, Flexner and colleagues (1963) injected intracerebrally several kinds of protein synthesis inhibitors in the hippocampi of stimuli-conditioned mice. They discovered that graded amounts of puromycin would impair the consolidation of recently acquired stimulus information. Unlike other pharmacological compounds used as control agents, Flexner el al's discovery unveiled that when peptide transfer is disrupted in the ribosome of hippocampal cells, memory consolidation is impaired. This important experiment revealed part of the molecular level of memory traces by manipulating a specific component and bringing about a differential effect in recollection. Ever since, different pharmacological and genetic manipulations have been used, and although we are far from distinguishing the neural mechanism underlying an event-specific memory trace, recent manipulations looking at differential effects in content-specific memory traces suggest that the search is promising. In a recent paper, for instance, Fellini and collaborators (2009) showed that NMDA receptors in CA3 are essential for pattern recognition tasks but not for spatial task, showing that controlled manipulations at the molecular level can illuminate the structure of content-specific memory traces.

Manipulations at the neurophysiological level have also shed light on the neural components of content-specific memory traces. The story begins with a widely cited study by Duncan (1949), in which he showed that electroconvulsive treatment could

impair the consolidation of recently acquired information—a finding that has been corroborated extensively (for a recent review, see Fraser et al., 2008). Unfortunately, the effects of electroconvulsive shocks are quite massive, and the precise reasons as to why they affect memory consolidation are unclear. Many neuroscientists hypothesize that electroconvulsive shocks interrupt protein synthesis temporarily, which in turn affects the polarization of the cell membranes blocking the transport of neurotransmitters (Fink, 1990). Fortunately, the depolarization component of the electroconvulsive shocks can now be isolated with the use of transcraneal magnetic stimulation (TMS), a non-invasive experimental technique in which a rapidly changing magnetic field sends off a weak electric current to a specific region in the cerebral cortex in order to produce a localized depolarization. Although seldom used in the context of memory given the difficulty of stimulating the medial temporal lobes, recent studies have explored the way in which depolarization affects memory retrieval. In a recent study, for example, Kohler and collaborators (2004) employed repetitive TMS to stimulate regions in the left inferior pre-frontal cortex (LIPFC), which were previously associated with successful encoding of the studied material (using the subsequent-memory paradigm, which I explain below). Participants who were stimulated in LIPFC showed higher accuracy for encoded words relative to both non-stimulated subjects and non-LIPFC stimulated subjects. Since it appears that repetitive TMS above 5 Hz transiently increases cortical excitability (Hallett, 2000)—an effect that parallels LTP—Kohler et al.'s study suggests that electric activity in the LIPFC is part of the mechanism underlying semantic memories.

Even more promising dissociations can be observed when we scale up a level. With the advent of non-invasive neuroimaging techniques, cognitive neuroscientists are

starting to identify brain regions that are differentially involved during content-specific memory retrieval. Two important lines of evidence are of particular interest here. The first line of evidence pertains to findings employing the subsequent memory paradigm (Wagner et al., 1998). In this paradigm, participants are asked to memorize content-specific stimuli (e.g., words, pictures, etc.) while in the MRI scanner. The recorded brain activity during encoding is then compared with the participant's responses for subsequently remembered versus forgotten stimuli. The use of the subsequent memory paradigm in cognitive neuroscience has revealed a network of interrelated brain regions, whose engagement plays a critical role during the consolidation of memory traces effectively leading to the recollection of particular episodes (Paller & Wagner, 2002). The other line of research pertains to one of the most consistent results in the research on the cognitive neuroscience of memory: remembering re-activates the sensory areas that were involved during the encoding of the retrieved material (Wheeler et al., 2000; Nyberg et al., 2000). The extent to which content-specific sensory cortices engaged during encoding are re-activated during retrieval has only recently started being studied. However, the results from these studies consistently show that visual information selectively re-activates visual cortices, auditory information selectively reactivates auditory cortices, and olfactory information selectively reactivates the olfactory cortices (Wheeler & Buckner, 2003; Gottfried et al., 2004; Woodruff et al., 2005; see Danker & Anderson, 2010, for a recent review)[7].

[7] Strictly speaking, neuroimaging techniques such as fMRI and ERP are detection rather than intervention techniques. However, as I am about to explain, when combined with intervention techniques, imaging methods can provide us with valuable causal information that we wouldn't have been able to gather otherwise.

One final line of evidence that speaks to the nature of the top mechanistic level of memory traces comes from neuropsychology. Departing from the observation that visual cortices were engaged during retrieval of visual memories, cognitive neuropsychologists David Rubin and Daniel Greenberg studied the nature of memory deficits associated with selective damage in the visual cortex (Rubin & Greenberg, 1998). They observed that, consistent with the sensory reactivation hypothesis, patients with damage in the visual cortex have trouble remembering visual details of previously encoded events, leading to what is now called visual memory-deficit amnesia (see Greenberg et al., 2005, for a review of 11 cases). Importantly, the psychological manifestation of the visual memory-deficit amnesia differs from the typical medial-temporal amnesia—such at H.M.'s—in that it only affects visual information; episodic information encoded non-visually or amodally (e.g., names) is spared. Brain lesions do not constitute direct manipulations, however, for it is hard to say whether a particular patient would have remembered a specific stimulus had she not suffered the brain lesion. A more controlled experiment would be called for. For instance, combining the subsequent memory paradigm and the TMS techniques reviewed above, cognitive neuroscientists could localize those brain regions preferentially engaged during the successful encoding of different stimuli (say, faces and houses), and then, during retrieval, they could selectively TMS each region. One would expect, therefore, that if the brain region that gets activated during successful encoding of a particular face, *x*, is part of the memory trace *x*, then by magnetically stimulating that very region one could control whether or not the subject successfully remembers having seen *x*.[8] As such, this would be direct evidence to the effect that there

[8] Along with cognitive neuroscientist David Pitcher (MIT), I am currently conducting this particular study, and we hope to get the results in a few months.

is an intermediary causal mechanism between the successful encoding of an event (*Ex*) (i.e., seeing face x), and its subsequent remembering (*Rx*)—or failing to remember (not-Rx)—the event. Such an intervention—to go back to the discussion of the previous section—would allow the cognitive neuroscientist to "switch" the causal path from B to C.

In sum, the few studies I just surveyed offer us a picture of the ways in which neuroscientists have manipulated, and *could* manipulate, memory traces at different levels. The putative mutual manipulability of memory traces requires that interventions done at one level affect the organization of the other levels. The fact that blocking NMDA receptors *but not* M2 receptors (Patterson et al., 1990) affects subsequent retrieval of content-specific memories, tells us that NMDA receptors, but not M2 receptors, are a part of that memory's trace. Likewise, if depolarizing the right occipital face area (rOFA) *but not* the right lateral occipital area (rLO) during recognition selectively impairs one's recognition of a particular face, this intervention would tell us that that the rOFA, *but not* the rLO, would be part of the memory trace of that face (see Pitcher et al., 2009). Taken together, the results of these studies are starting to give us a picture of the neural underpinnings of memory traces that resembles the multi-level structure of the causal mechanisms involved in cyanide poisoning. Suitable interventions at the right level of the neural mechanisms composing memory traces may give us the differential effects in recollection that the anti-realist was unable to explain.

Let me finish this section with a clarification. At this point, someone could object to my description of memory traces arguing that the characterization I just offered, and the empirical evidence I reviewed, applies to memory systems and processes—such as

encoding and retrieval—rather than memory traces per se. This is a valid objection, but only if one accepts a distinction between the mechanisms of encoding, storing and retrieval as they apply to a particular memory of an event, and the memory trace of that particular event. However, I reject this distinction. I believe memory processes and memory traces are essentially intertwined. As I will explain at length in chapters 4 and 5, memory traces are tantamount to dispositional properties of multi-level neural mechanisms to (more or less) reenact the pattern of activation they were in during encoding at the time of retrieval. This blurs the distinction between the processes of encoding and retrieval of a particular experience, on the one hand, and the memory trace of such an experience, on the other. The distinction between the two is, at most, nominal. Still, I believe the notion of memory trace is useful insofar as it allows us to refer to the causal mechanism responsible for the retrieval of individual memories.

5. Experimental realism and the reality of memory traces.

At the beginning of this chapter I made a distinction between two ways of understanding the hypothetical status of memory traces. On the one hand, there was the philosopher's way, according to which memory traces were a theory-independent hypothesis to be verifiable conceptually. On the other hand, there was the psychologist's way, according to which memory traces were a theory-dependent hypothesis to be verifiable empirically. Next, I argued that the way in which philosophers have tried to settle on an answer for the theory-independent question is by wondering whether or not memory traces are required in order to furnish satisfactory causal explanations of remembering. Realists about memory traces claim that we should infer the existence of

causally mediating mechanisms during recollection from the claim that their postulation constitutes the best causal explanation for the phenomenon of remembering. Conversely, anti-realists about memory traces argue that causal explanations that do not postulate the existence of memory traces are equally good causal explanations for the phenomenon of remembering. However, I tried to cast doubt upon this last assertion showing how the anti-realist is incapable of explaining differential effects in recollection without alluding to intermediary causal mechanisms. As a result, I suggested that a successful causal explanation of remembering ought to involve reference, not only to the remembered event, but also to intermediary memory traces. As I suggested, memory traces become explanatorily indispensable precisely because their existence would be required to account for certain differential effects in recollection. The specific way in which memory traces are causally explanatory was then justified in terms of a manipulability theory of causation, and the exact nature of such manipulations was then described within a mechanistic framework. Accordingly, memory traces were defined as multi-level neural mechanisms of the kind studied by psychologists and neuroscientists. The two ways of understanding the hypothetical status of memory traces have thus collapsed.

Still, the philosopher may be unhappy, for even if the postulation of memory traces is required for a successful causal explanation of remembering, one can still wonder whether that is enough reason to infer that they exist. Framed as such, the question about our justification for accepting the existence of memory traces becomes an instance of the more general question about the justification for the existence of theoretical posits in the philosophy of science. Thankfully, there is a prominent view that indicates that we may be justified in inferring the existence of memory traces on the basis

of their indispensability in causal explanations of remembering. This view, initially presented by Ian Hacking (1983) and Nancy Cartwright (1983), goes under the rubric "experimental realism", although sometimes it is also known as "entity" or "instrumental realism". Experimental realism is better understood as a mid-point between scientific realism and anti-realism. It differs from scientific realism in that it is *not* committed to the existence of those theoretical entities that are postulated merely on the basis that they are required by the best theories we have. Experimental realism is agnostic as to whether or not *fundamental* theories are true—that is, theories containing general explanatory claims aimed at abstracting away from specific experimental circumstances to include general cases. Instead, experimental realism advocates a more austere commitment only to those entities that feature in the more local or *phenomenological* generalizations employed in causal explanations of circumscribed cases, predominantly in the context of experimental manipulations. As a result, experimental realism differs from scientific anti-realism insofar as it does accept the existence of *some* unobservable entities, to wit those that are required by the phenomenological generalizations used in successful causal explanations (Cartwright 1983; see also Clarke, 2001; Hoefer, 2008).

The explanatory indispensability of memory traces I have argued for agrees with the spirit of experimental realism. After all, when suggesting that memory traces ought to be treated as multi-level neural mechanisms I remain agnostic as to the structure of the general theory of memory that would ultimately incorporate them. Currently, in the cognitive neuroscience of memory literature, there are at least two general views about memory encoding and retrieval that offer two different ways of understanding memory traces. On the one hand, there is the traditional consolidation view according to which the

hippocampus is only required during memory encoding, because once the memory trace is consolidated in the sensory cortices, the frontal lobe is capable of retrieving the memory on its own (Frankland & Bontempis, 2005). On the other hand, there is the multiple trace theory, according to which the hippocampus is still required to retrieve the memory trace, so that the frontal cortex cannot do it on its own (Nadel & Moscovitch, 1998). Both of these views are general theories about the nature of memory encoding and retrieval, and it is still an open question as to which one of them is true. However, when it comes to the more local explanations of differential effects during recollection due to well-controlled experimental manipulations, the existence of memory traces is justified, even if ultimately both the traditional as well as the multiple trace theory have a story to tell as to how to accommodate the results. The hope is, of course, that neuroscientists will keep on perfecting experimental interventions such that the different putative properties of memory traces are properly manipulated, so that eventually one theory would trump the contenders. But the question as to whether or not there are memory traces is independent of this ever occurring.[9]

[9] Needless to say, experimental realism—just as any other view on the realism/anti-realism debate in the philosophy of science—has been widely criticized (e.g., Hitchcock, 1992; Resnik, 1994; Reiner & Pierson, 1995). Nonetheless, many believe that it still remains a strong contender in the dispute (e.g., Clarke, 2001; Suarez, 2008). Thus, if is good enough for them, it is good enough for me. Any further defense of experimental realism as a tenable position in the realism/anti-realism debate is beyond the limits of this chapter.

3. Is Remembering a Propositional Attitude?

When I was younger, I could remember anything, whether it happened or not.

Mark Twain

1. Introduction

Remembering is a mental event. Therefore, it is also an intentional event: our memories are *about* something. That which our memories are about we call *intentional objects*. They are so-called in order to avoid the ambiguity with mere existing or actual objects, for the objects of our memories need not exist when we remember; in fact, they usually have ceased to exist, for what we usually remember are past events. In addition, memories have *intentional contents*. Since intentional objects can be present to the mind in many ways—even in ways that may not accurately correspond to how they are—we need a notion for their distinctive mode of presentation. The mode in which an intentional object is mentally presented is the intentional content of a mental state (Crane, 2009).

Traditionally, the intentionality of remembering has been captured by treating it as a propositional attitude. The same goes for most memory experiences (e.g., remembering, recalling, reminiscing, etc.) that can be expressed, without much linguistic maneuvering, in the canonical form "S remember(s) that p". I say *most* because there are some memory experiences that seem to resist such treatment, as when we say "Mike remembers how to ride a horse". In this case, it seems as though we are saying, of Mike,

not that he has a particular attitude toward a proposition, but rather toward a particular disposition to act (i.e., if he was to try to ride a horse he'd be able to do it). However, occurrent memory experiences, of which remembering is paradigmatic, do seem amenable to be treated as propositional attitudes. Indeed, so-called "factual" or "propositional memories" (Bernecker, 2009) have been conventionally treated as propositional attitudes, irrespective of whether they are about experienced (e.g., "I remember that you were wearing a hat at my party") or non-experienced facts (e.g., "I remember that the Declaration of Independence was signed before the French Revolution began"). Since the received view sees propositional attitudes as relations between a subject and a proposition, most philosophers of mind working on memory are inclined to treat remembering along the same lines. Specific theoretical variations notwithstanding, the assumption that remembering is a propositional attitude has seldom been challenged, largely because it allegedly finds support in the logical form of memory reports and the apparent propositional nature of their its intentional contents.

In this chapter I challenge the assumption that remembering is a propositional attitude. Although the bulk of my argument pertains to memory of experienced events—what psychologists call 'episodic autobiographical memory'—it is possible that some aspects of it translate to other kinds of memory, such as semantic or implicit memory. My main focus will be, then, arguing that factual episodic remembering should not to be understood as a relation between a subject and a proposition. My argument is two-tiered. In part 2, I offer an argument against the relational account of propositional attitudes *in general*, while stressing the way in which it applies the case of remembering *in particular*. The upshot is that the relational view—possibly for all propositional attitudes

in general, but certainly for remembering in particular—is either empirically improbable or explanatorily useless. In part 3, I explore an argument in favor of remembering as a propositional attitude that is independent of the general relational view about propositional attitudes. This is an argument in support of the view that the intentional content of our episodic autobiographical memories is propositional. In turn, I offer two counter-arguments against such a view. I suggest instead that the intentional content of our episodic autobiographical memories is non-propositional. Finally, in part 4, I explore some of the consequences of giving up the notion of remembering as a propositional attitude, and I present them in the form of a challenge—a challenge that I try to meet in the next chapter.

2. Against remembering as relational

According to the received view, remembering is a propositional attitude. Propositional attitudes are typically viewed as mental states that involve both a proposition and an attitude toward that proposition, so they are normally treated relationally (Schroeder, 2006). It is also typically assumed that propositional attitudes are the mental states we refer to when we use propositional attitude reports. In its most general form, propositional attitude reports are expressed by sentences composed by a subject, an intentional transitive verb, and an embedded clause that the verb takes as its complement. Examples of these are:

(1.1) Mary believes that Jack is the killer.

(1.1) Miguel and Jose think that you shouldn't be telling that joke in public.

(1.3) My mother hopes that I come home for Thanksgiving.

(1.4) I remember that you were wearing a hat at my party.

Additionally, it is customary to regiment propositional attitude reports in the following fashion:

(2) (∃*S*) (∃*p*) (∀R) (R(*S*,*p*))

where '*S*' refers to a subject, '*p*' refers to whatever the referent of the sentential complement clause may be, and 'R' to the relevant intentional relation between them (e.g., Fodor, 1978/1981, 178; Schiffer, 1992, 491).

This relational view, according to which propositional attitudes are relational mental states between subjects and propositions, is thus motivated by two interrelated assumptions. The first assumption—call it (a)—is that (2) is in fact the logical form of memory reports such as (1.1 – 1.4) (but see Moltmann, 2003, for a different view). The second assumption—call it (b)—is that you can infer the nature of the mental state expressed by (1.1 – 1.4) via reading off its underlying logical structure—namely (2)—and accepting the existence of the objects picked by the bound variables upon which (2) existentially quantifies. Unsurprisingly, what these objects may be is a matter of controversy. Leaving aside the (underappreciated) problem of how to understand '*S*', it seems to me that there are two general strategies when it comes to cashing out the ontological value of '*p*'. On the one hand, there are those who think that the subject is related to a non-mental entity, which is usually—albeit not always (see below)—conceived as a truth-evaluable proposition. Call this approach *non-mental relationalism*. On the other hand, there are those who think of it as a mental entity tantamount to a mental representation, particularly a sentence-like formula in the language of thought. Call this view *mental relationalism*.

I think the relational view is problematic on two grounds. First, the assumptions, (a) and (b), which motivate the relational view are wrong or, at best, insufficient. And second, even if these assumptions were right, both non-mental and mental relationalism are unsatisfactory accounts of the nature of some mental states, of which remembering is but a case. At least for the case of remembering, I submit, non-mental relationalism faces what I call *the problem of traction*, which roughly consists in being unable to explain how our memories can play a causal role in our behavior. Although mental relationalism could in principle avoid this difficulty, it faces what I call *the problem of evidence*: it has no empirical support. Let me elaborate.

Although both (a) and (b) are, I think, highly contentious claims, they are seldom argued for. When talking about "belief", for instance, Schiffer simply assumes that it refers to "the relation expressed by "believes" in a sentence of the form "x believes that S' (1992, 500). Similarly, Fodor claims that 'believes' "looks like a two-place relation, and it would be nice if our theory of belief permitted us to save appearances" (1978/1981, 178). Thus, if we were to import this argumentative line to the case of remembering, the upshot would be that since "S remembers that p" looks relational, then remembers must be a two-place relation as well. Certainly this would permit us to save appearances. Unfortunately, it isn't clear what are the appearances we want to save—let alone why. Do we want to save the appearance that, in our natural language, remembers looks like a two-place relation? If that is so, then I wonder what "our" is supposed to range over. Today, there are between 5,000 and 8,000 languages in the world (Evans & Levinson, 2009). Each one of them constitutes a natural language. It would be a miracle if all natural languages would have a lexicalized verb not only sharing the same semantic field

as the English verb "to remember", but also taking as complement a sentential clause. Although less that 10% of the languages spoken right now in the world have been decently documented, we already find counterexamples. In Dalabon, a gunwinyguan language of Central Arnhem Land, Australia, there are simply no lexical verbs dedicated to remembering. With basic distinctions between "fleeting" and "enduring" mental states, and through the use of tenses and aspectual transformation, speakers of Dalabon manage to talk about events that happened to them in their past that have endured, and that are now fleetingly present to their minds. But no single lexical component is used for "remembering" alone, as the words employed to talk about memories can be also included in diverse grammatical constructions to mean different things, like "realize", "attend to", "think" and "decide" (Evans, 2007). Moreover, there seems to be evidence that some languages do not even make distinctions between verbs and their direct complements. In Straits Salish, an almost extinct language spoken by the Salish people in the American pacific northwest, there is only one major class of lexical item functioning as predicate, as is the case with intransitive verbs in English (e.g., "It rains". Jelinek, 1995) In neither of these languages do their approximate cognates of "remembers" look relational—or at least they do not look relational in the same way in which "remembers" looks relational in English.

Now, why aren't *these* the appearances we wanted to save? A possible answer is to say that the differences among languages do not really matter, for they are all translatable into languages for which the canonical form of propositional attitude reports apply. But this response only pushes the argument one step back, as we can wonder why we want to translate the expressions in natural languages for which the canonical form

does not apply to expressions in natural languages for which it does. Without a principled reason to prefer the latter versus the former, I don't see why we should favor some linguistic appearances over the others. Another possibility is to claim that semantics is language-bound, so the question as to whether or not propositional attitude reports refer to relational mental states only arises in those languages in which propositional attitude reports are expressed in conformity with their canonical form. This answer would be just fine, I contend, if philosophers were willing to stop at the semantic level, and refrain from drawing metaphysical conclusions from their linguistic observations. Unfortunately, that is not the case, as some of them—predominantly Fodor (1975; 1985)—go the extra mile. Obviously, though, if we want to hold that propositional attitudes are relational as a matter of metaphysical necessity, we had better not base our argument in the contingent fact that we speak English.

Another possibility is to say that the appearance we want to preserve is not on the surface but at the deep grammar level. Thus, even if a language lacks a lexicalized verb for "remembering", it may be possible that remembering-like constructions can conform, at the deep grammar level, to the canonical form of propositional attitude reports. This strategy is problematic too, for even at the deep grammar level, the distinction between complement, relative, and adverbial clauses isn't always clear-cut. For instance, sometimes determining whether a subordinate clause pattern conforms to one or another structure may be, more or less, a matter of taste. "I remember when I used to play" seems to take as a complement an adverbial clause, but for certain purposes could be taken as a relative clause with an elided head noun, e.g., "I remember [the days] when I used to play" (Evans, personal communication). Forcing all construction patterns to look like

nominal phrases taking as complements sentential clauses of the form "S remembers that p" may look like an attempt to make the data fit the theory rather than the other way around. Finally, if the appearance we want to preserve isn't grammatical at all, but merely conceptual, the evidence looks even grimmer. Recent attempts to document the conceptual structures underlying expressions of memory experiences across cultures has revealed that REMEMBERING is not a universal basic concept, but a construct that, depending on the culture, is built upon more basic concepts like KNOW, THINK and BEFORE (Wierzbicka, 2007). In sum, not only there is evidence that speaks against the universality of the concept of remembering, but there is also reason to believe that, at least in some languages, memory reports do not conform—neither on the surface nor at the deep grammar level—with their alleged canonical form, or in ways that take as complements referential 'that'-clauses. This lack of uniformity in their linguistic expression, I believe, should at least make us wonder whether (2) is indeed *the* logical form of memory reports.

Claim (b) isn't less contentious. Even if we grant that, for whatever extraordinary means, all languages could render their own expressions of memory experiences in the canonical form of propositional attitude reports, and even if we grant that the logical form of such sentences is conveyed by an instance of the formula (2), we still don't have to be committed to the claim that the intentional verb (i.e., 'remembers') denotes a relation to a proposition[10]. The assumption underwriting claim (b) is that by reading (2) we have reason to believe that the entities referred to by '*p*' (i.e. the 'that'-clause) exist because '*p*' is bound by an existential quantifier. The argument for this assumption goes roughly

[10] Much of what I am about to say probably applies to other propositional attitudes in addition to "S remembers that p". However, for the purposes of the present chapter, I will confine my discussion to the case of remembering.

like this: the logical form of propositional attitude ascriptions tells us that (P1)'that'-clauses are referential. We know that (P2) referential terms can be bound by existential quantifiers, and that (P3) if a referential expression is bound by an existential quantifier, we have reason to believe that the entity referred by it exists. Thus, (P4) since 'that'-clauses are bound by existential quantifiers, then (C) the referents of 'that'-clauses exist.

Unfortunately, premises (P1) and (P3) may not be true. In the last two decades we have seen a surge of semantic puzzles pertaining to 'that'-clauses in propositional attitude ascriptions (e.g., failures of substitution, failures of existential generalization, Kripke/Bach's "Paderewski" cases, etc.), so many philosophers are starting to consider that they may not be referential at all—or, at the very least, that they may not refer to propositions (Bach, 1997; Moltmann; 2003; Hofweber, 2006). The truth of (P1) is, therefore, at least questionable. In addition (P3) seems problematic as well. For the conditional to be true, one needs to be committed to an ontologically loaded reading of existential generalization, very much in the spirit of Quine (1948).[11] But existential quantifiers can be read as ontologically innocent, for instance in line with what Hofweber calls 'an internalist view', that is, as logical devices that allow us to increase expressive power in order to talk about infinitary disjunctions of single instances (Hofweber, 2006). Under that reading, even if 'that'-clauses turn out to be referential after all, still we wouldn't have to be committed to the existence of whatever they purportedly refer to. In sum, since 'that'-clauses may not be referential, and since not all referential expressions

[11] As I have argued elsewhere (but see also Balaguer, 1998), the argument from existential quantification to the reality of propositions is based on Quine's criterion of ontological commitment plus an "intentional" reading of the Quine-Putnam indispensability thesis (see De Brigard, 2007; for the view I am arguing against, see Fodor 1978/1981). Of course, Quine wanted to dispense with intentional talk, so he wouldn't have pushed for realism about propositional attitudes. But Fodor (and many others) have argued against Quine on this point, suggesting that intentional talk is neither reducible nor eliminable (Fodor, 1974). Accepting this line of argument gives us the indispensability we require to apply Quine's criterion of ontological commitment.

bounded by existential quantifiers need to be treated as real, there is no a priori reason to think that the referent of 'that'-clauses exist.

To be sure, these difficulties haven't gone unnoticed by relationalists. Two of the most promising alternatives are the so-called 'hidden-indexical' theories (Schiffer, 1987; Crimmins and Perry, 1989) and Richard's 'quasi-Russellian' account (1990). According to the former, attitude-verbs like "remembers" have an unarticulated indexical as a third component. Thus, the logical form of sentences like (1.1 – 1.4) is not (2) but rather

(3) $(\exists m)(\mathrm{M}m \wedge (\forall \mathrm{R})(\exists S)(\exists p)(\mathrm{R}(S,p,m))$

where '*S*' refers to the subject, '*p*' to the remembered proposition, '*m*' to a mode of presentation of type M, and 'R' to the three-place relation of remembering which *S* bears under *m* towards *p*. By introducing a hidden-indexical referencing a particular mode of presentation (which, I take it, could be construed as an intentional content), this view manages to avoid inconveniences having to do with substitution failures, anaphoric relations, and quantification. Alternatively, Richard's quasi-Russellian view introduces the notion of 'Russellian Annotated Matrix' or RAM. There are two kinds of RAM: a public-language RAM and a language-of-thought RAM. The former is the ordered pair consisting of a sentence and a Russellian proposition expressed by an utterance of that sentence in a context, while the latter is the ordered pair consisting in a formula in the language-of-thought and the Russellian proposition that interprets it (Matthews, 2007: 107). So, according to this view, in "John remembers that Mary is home", the sentential clause "that Mary is home" denotes the public language RAM ⟨⟨'is home', being home⟩, ⟨'Mary', Mary⟩⟩ which *represents* the subject's language-of-thought RAM referring to the fact that Mary is home (Richard, 1990: 13ff). Since the public utterance can shift

contexts, the contextual sensitivity of the ascription is built into the formula that relates the two kinds of RAMs. Thus, according to this view, the correct regimentation of propositional attitude reports would be:

(4) $(\exists f)(\mathrm{F}f \wedge (\forall \mathrm{R})(\exists S)(\exists \mathcal{R})(\mathrm{R}(S,\mathcal{R}_c,f))$

where S refers to the subject, $\mathcal{R}_c$ to the RAM-in-context-c determined by the 'that'-clause, 'f' to a context-sensitive function of representation of type F, and 'R' to the remembering relation between S and $\mathcal{R}_c$ under f (Matthews, 2007: 108).

Despite the apparent success of these sorts of views at handling the aforementioned semantic puzzles, for the purposes of understanding the nature of remembering qua mental state, they pose more questions than they answer. For one, both the hidden-indexical and Richard's quasi-Russellian accounts take 'remembering' as a three-place relation, not as a two-place relation as (a) assumes. Second, by introducing a third component to that relation, these accounts double our metaphysical qualms, for now we have to ponder its nature and how it relates to the other two. But most importantly, these accounts, just as the traditional non-mental relational views, suffer from what I call (loosely following Matthews, 2007: 105) *the traction problem*: any relational view of remembering (or any other propositional attitude, for that matter) that posits as a relatum some sort of abstract entity—be it a proposition, a set of possible worlds, a RAM, or even a presently existing past event—is going to have a hard time explaining how such an entity manages to be causally efficacious. And if we want our propositional attitudes to figure in our psychological explanations, we better have an account as to how they manage to produce behavior. [12]

[12] Notice that this is not only a problem about how propositions manage to cause psychological states *in virtue of* their content. That's certainly a thorny issue. But the concern I am rising here is even more basic.

Richard (1990) is aware of this problem, though, as the role of *f* in (4) can be seen as an attempt to get around it. Recall that he distinguishes two different objects of propositional attitudes: a semantic object (i.e. the public language RAM) and a psychological object (i.e. the language-of-thought RAM). Accordingly, the semantic object manages to resolve the semantic inconveniences, while the psychological object—which is ultimately a formula in the language-of-thought—does the causal heavy lifting (Larson & Segal, 1995, advocate for a similar distinction). Now, the semantic and the psychological objects are related by *f*, a representational function that maps the semantic object onto the psychological object in different contexts. The problem is that the nature of this relationship not only is vague, but it is also potentially damaging for the relationalist enterprise. Here is why. If *f* is a mapping function, then it could be either isomorphic or non-isomorphic. Suppose that *f* is isomorphic. If so, then the structure of the semantic object mirrors the structure of the psychological object. This means that if a particular semantic object, *a*, means *X* in context *c*, but *Y* in context *d*, then the same contextual sensitivity must be present in the psychological object, *b*, that *a* maps onto. But psychological objects are not supposed to be context sensitive in the same way in which semantic objects are, in order to guarantee that the subject can think of the same thing in different contexts. So there is at least one structural property that semantic objects have that their corresponding psychological objects do not.[13] Which means that *f* is non-isomorphic. But if *f* is non-isomorphic, then propositional attitudes need not be relational, for the logical structure that invite us to think of them as relational pertains to

---

If one posits propositions as part of the explanation for the causes of our psychological states and behaviors, one better be prepared to explain how propositions can cause anything at all.

[13] Of course, there are many more structural properties that would work equally well for this line of argument, e.g., phonetic constructions, syntactic modifications, etc. (Larson & Ludlow, 1993).

the semantic and not the psychological object and we have no reason to believe that the structure of the latter tracks the structure of the former.

Surely one can avoid this problem by getting rid of *f* altogether, and the most natural way of doing it is assuming that the psychological and the semantic objects are one and the same; this is basically the approach of mental relationalism. According to mental relationalism, the object of a propositional attitude is a mental representation whose content is that of the sentential 'that'-clause. In general, in order to avoid the problem of traction, mental relationalism subscribes to a representational/computational theory of mind. The most well known account of propositional attitudes in these terms is the Language of Thought hypothesis (LOT). Roughly, LOT suggests that '*p*' in (2) refers to a mental representation in a symbolic/computational system, which is built out of semantically simpler representations in virtue of the system's combinatorial syntax, and whose computations are defined over the syntactic or formal structure of their symbols. According to this view, a memory report such as (1.4) should be understood as a computational relation—whatever that means—between a subject (in this example, me) and a mental representation in my internal LOT, which is implemented in my brain, and whose meaning is that you were wearing a hat at my party. Now, suppose that I also learned that the only person with a hat at my party was responsible for the stain in the rug, and as a result I ask you to pay for the damages. Following LOT, the explanation would be, roughly, that the mental symbols that played the role of referring to you in my remembering are the same ones that were used for the existential instantiation of my belief that the only person that was wearing a hat was responsible for the stain. In turn, this explains why my conclusion has you as the subject of my desire that you pay for the

damage. And this all occurs in such a harmonious manner because the computations driving my reasoning are sensitive to the syntactic form of my mental representations (which, ultimately, is some form of brain activity), and their syntax preserves their semantics.

Unfortunately, years of research in philosophy of mind and cognitive science have shown that LOT faces what can be called *the problem of evidence:* it does not have much empirical support. LOT was originally proposed as a hypothesis to the best explanation for, basically, two sets of reasons. First, it was allegedly the only view that could explain language and concept acquisition, reasoning by hypothesis formation, and the compositionality, productivity, and systematicity of thought. Second, it was supposed to receive empirical support from behavioral evidence coming from psycholinguistics, visual processing research and theories of reasoning. Unfortunately, both sets of reasons have been severely undercut during the last two decades. Regarding the latter, there are many contemporary views of language processing that can explain the same phenomena LOT was set up to explain, without using explicit sentential representations (e.g, Barsalou, 1999; Chalmers, 1996). Likewise, mental-model theories (Johnson-Laird, 2010) have proven to be at least as successful as LOT-like views in explaining reasoning, and connectionist models of vision are actually better at handling visual phenomena LOT proved unable to (for careful reviews, Matthews 1989; 2007: ch. 3). Regarding the first set of reasons, things don't look better. Empirical approaches to language and concept acquisition and processing are now equipped to deal with the problems LOT was supposed to better handle, and the claim that non-sententialist representational systems cannot handle compositional and systematic semantics is simply not true. Using

hypotheses to the best explanation in support of a realist stance regarding certain sorts of entities—in this case, the referent of 'that'-clauses in expressions of propositional attitudes—is always a risky business, as such hypotheses are always relative to the set of explanations available to us at a particular time (van Fraassen, 1980). Since the contents of these sets can fluctuate with time, it is always possible that a hypothesis, H, is the best explanation of certain phenomena at $t_1$but not at t_2. And if the reality of the theoretical entities posited by H was supported by H's being the best hypothesis at t_1, its demise at t_2 certainly undercuts our reasons to believe in them—even in the absence of a competing view! Defenders of LOT wrongly argue that in the absence of another theory that can explain the same phenomena that LOT explains, LOT is still "the only game in town". This is false on two grounds. First, a theory can be falsified even if no alternative theory with the same explanatory scope is put forth, and second, it is not clear that the phenomena LOT is supposed to account for need to be explained by the *same* view. LOT purports to be a theory of thought, but 'thought' may not name a unified set of phenomena; maybe thought is not even a natural kind. As such, there is no reason to think that a single theory needs to explain thought. After all, a false unifying theory isn't better than several discrete ones with a better chance at being true—or, for that matter, than no theory at all.

Let's take stock. The received view on remembering has it as a propositional attitude. Traditionally, propositional attitudes have been treated as relations between subjects and semantically evaluable entities. This relational view is supported by the claim that the logical form of expressions of propositional attitudes is relational, and that such logical form reflects reality. I argued against that relational view by pointing out,

first, that the logical form of expressions of remembering may not be relational and, second, that even if it were, there is no good reason to believe that it reflects reality. In addition, I argued that the two preferred views of propositional attitudes—non-mental and mental relationalism—would face problems when it comes to explaining remembering (and, presumably, other propositional attitudes). On the one hand, non-mental relationalism faces the traction problem, as it isn't clear how a non-mental truth-evaluable entity can play a causal role in the production of behavior. If I want to explain why my remembering that you were wearing a hat at my party is causally responsible for my asking you to pay for the carpet, I need an account that can tell me how the content of my memory played a role in my psychological process. On the other hand, I argued that although mental relationalism could, in principle, avert the traction problem, it faces a problem of its own: lack of evidence. Without empirical support, there is no reason to believe that when we remember we are psychologically related to a sentence in an internal code.

3. Against remembering as propositional

The above arguments would only affect the view of remembering as a propositional attitude if one takes an ontologically loaded reading of the relational structure of memory reports. However, as I mentioned above, one need not draw any ontological consequences out of this semantic fact. Remembering can still be seen as a propositional attitude, from a semantic point of view, merely because the intentional content of our (episodic autobiographical) memories is propositional. This observation goes beyond mere linguistic structures, and is grounded on the apparent fact the

intentional content of our memories is truth-bearing; that is, it can be either true or false. Indeed, the claim that the content of our memories is truth-bearing receives further support from one of the most cherished assumptions in the philosophy of memory literature: the so-called *factivity constraint* (Bernecker 2009, 137). According to the factivity constraint, "to remember" belongs to the class of verbs known as 'factive verbs'. As such, an utterance or sentence such as 'S remembers that p' is true only if 'p' is true. And if so, then 'p' must denote the sort of thing that can be true. Since propositions are truth-bearers par excellence, it follows that the content of our memories is propositional.

Unfortunately, as I plan to show, there are two problems with this argumentative line. On the one hand, I believe that the factivity constraint is false. On the other, I think that memorial contents, unlike propositional contents, admit of degrees of correctness; they aren't solely true (or false), but rather more or less correct. So let me begin with an examination of the factivity constraint.

Although seldom argued for, the factivity constraint is usually introduced after a cursory exploration of a handful of readily available examples of memory claims. Consider the statements:

(5.1) I remember that Mary was wearing a green hat.

(5.2) I remember opening the door with my own key.

(5.3) Jose remembers that the car in front of his was scratched.

According to the factivity constraint, if (5.1 – 5.3) individually are true, then it follows that Mary was wearing a green hat, that I opened my door with my own key, and that the car in front of Jose's was, indeed, scratched. The truth of a memory report implies the truth of that which is reported as remembered. If I utter (5.1) and it turns out that Mary

wasn't wearing a green hat, then I must say that I wasn't really remembering. At most, I might say that I thought I was remembering, or that I seemed to remember, not that I was actually remembering. One can only remember things or events that were the case.

Virtually every endorser of the factivity constraint assumes it with no other argument besides the (often unsupported) claim that it is implied by the way in which competent speakers use the word "remembering" (e.g., Audi, 1998; Malcolm, 1963; Shoemaker, 1972.). Only a handful of philosophers have tried to provide some support for what otherwise is thought of as an obvious claim. One line of argument is to claim that the conjunction of a memory claim with the negation of its embedded clause is contradictory. According to this line of argument, if someone is to utter

(6) I remember I that was drinking tequila but I wasn't drinking tequila.

then she would be contradicting herself. But (6) isn't really a contradiction. It is only incoherent in the same way in which Moore's famous "It is raining outside but I don't believe it" is incoherent (Hazlett, 2010). After all, it would be a mistake to think that no competent user of the verb "to remember" can rationally hold (6) true. At most, (6) is only *pragmatically* incoherent, and the incoherence appears solely when the conditions, under which the claims before and after the "but" are evaluated, remain fixed. Thus, if we are to evaluate both claims according to a simple description of the recounted event, incoherence ensues. But the evaluative conditions between claims can easily shift. Consider a case in which a true Mexican charrito, known for his exquisite taste for tequila, parties with a cheap bottle of Jose Cuervo. In this case the first claim is to be evaluated as a mere description of the event, whereas the second claim is to be evaluated in terms of different standards of what constitutes tequila. That is why he can coherently

say that what he remembers drinking was tequila but it wasn't really tequila (see Bernecker, 2010).

Another possible argument in favor of the factivity constraint is grammatical, and it consists in applying Vendler's citerion to the verb "to remember". According to Vendler (1972), one could distinguish between factive and non-factive verbs in that the former, but not the latter, can be transformed into wh-clauses. Thus, "to remember" is factive because it can take the form of "Jose remembers where the car in front of his was scratched", "Jose remembers when the car in front of his was scratched", etc. But what is really the scope of this argument? First, albeit grammatically correct, it seems to make little psychological sense. It may very well be that Jose does not really remember where the scratch was, let alone when it happened. There seem to be cases in which the wh-clauses, produced by vendlerizing remembering statements involve information about the remembered event, which need not have been considered in the original statement. Should we say that these aren't genuine cases of remembering, then? If we say that they are, then the Vendler criterion isn't really that useful. At most it indicates something interesting about the grammar of the English verb "to remember" that has no bearing on the truth of its contents. But if we say that they are not genuine cases of remembering, precisely because there is some information "missing", then we are unduly constraining our cases of genuine remembering to cases in which an enormous amount of information about a particular event needs to be brought to mind. Almost none of our memories would count as genuine memories then.

A second reason to be suspicious of this strategy is that it is hard to apply neatly to memory statements about events that haven't occurred or events that are atemporal in nature. Consider:

(7.1) I remember that I am going to see you next week.

(7.2) I remember that the number of planets is 9.[14]

In these cases, the following wh-clauses derived from (6) and (7) sound odd, if not grammatically, at least pragmatically:

(7.1*) I remember when I will be seeing you next week.

(7.1**) I remember what I will be seeing you next week.

(7.2*) I remember what the number of planets is 9.

(7.2**) I remember when the number of planets is 9.

A possibility is to try to translate these statements into remembering-statements about events that happened in the past. Thus, we could say that (7.1) and (7.2) really mean:

(8.1) I remember that I made an appointment with you for next week.

(8.2) I remember having learned that the number of planets is 9.

But as Munsat (1966) showed, this strategy won't do. When I remember that I will be seeing you next week I need not bring to mind anything at all regarding the way in which I encoded the information about our future meeting, not even images of me writing it down in my personal schedule. I may remember that I will be seeing you next week without remembering that I made such an appointment. Likewise, many a times we remember lots of facts, like the fact that the number of planets is 9, without remembering

[14] I am aware that (7.2), unlike all other examples I've discussed, does not refer to an autobiographical episodic memory but to a semantic memory. Also, (7.1) is a case of prospective memory, which arguably (although less clearly) may not count as a bona fide case of episodic memory. However, since the point I am trying to make here is against the factivity constraint, which in turn supposedly applies to memory in general, that my examples aren't confined to episodic autobiographical memory is beside the point.

when or how we learn them. To say that I remember that the number of planets is 9 does not imply that I also remember having learned it. I may have forgotten learning about it while I still remember the fact.

Perhaps the most obvious argument against the factivity constraint is the simple fact that competent speakers just don't abide by it when they use the word "remembering" (see Hazlett, 2010, for some evidence to this effect). Talk of false memories, for instance, is nowadays ubiquitous. Most people feel comfortable using the word "remembering" when referring to things that did not happen, or things that did not happen exactly the way in which they remember them happening. Are people just misapplying the word "remembering" across the board? If we follow the tradition here, its seems as though the answer philosophers have given is 'yes', as they have quickly dismissed the concern about false memories by distinguishing between *ostensive* and *veridical* remembering (Shoemaker, 1972). According to this distinction, ostensive memory is only "seeming to remember", whereas veridical remembering is, well, *true* remembering, i.e., the mental state of recollecting what in fact was the case. The thought is, therefore, that when we use the word "remembering" when referring to events that did not happen, or events that did not happen the way in which we remember them, we are speaking loosely, for we should have used instead the locution "seeming to remember". But I am not sure this strategy is warranted. Notice that the locution "seeming" has, at least, two possible senses (Schwitzgebel, 2008): an epistemic sense and a phenomenological sense. In the epistemic sense, we use "it appears" or "it seems" to indicate hesitation or uncertainty, as when I say, for instance, that I seem to remember having left the keys inside my car. False and distorted memories just don't appear to us in

a way that makes us hesitant about them being memories at all. In contrast, in the phenomenological sense we use the locution "it appears" or "it seems" to indicate the way in which a particular mental content presents itself to our consciousness, as when we say, looking at the Muller-Lyer illusion, that one line appears longer than the other. In this case we express no hesitation but a mere phenomenological report. The problem is that, when it comes to the phenomenological sense of "seeming to remember", distorted and veridical memories are indistinguishable. The distinction between seeming to remember and actually remembering only makes sense from the point of view of epistemology, but this is because the philosopher has already confined his notion of remembering to veridical memories—a decision that is not grounded in the way competent speakers use the word "remembering".

This last point actually dovetails with the second reason why I think memorial contents may not be propositional. Suppose you entertain a mental state you think is a memory but someone gives you a piece of information that reveals inconsistencies between the way the content of the memory is present to you, and the way the event—at least according to that person—actually happened. Sometimes, when this occurs, one of us just admits that he was wrong and that the event occurred as the other person reported it. But most times the disagreements aren't settled in terms of right and wrong. Many times people simply agree that they remember events differently, with more or less detail, or more or less accurately. There is usually no disagreement in telling whether a version is right and another one is wrong, but when it comes to correct versions, there may be disagreement as to which one is more or less accurate. Accuracy admits of degrees. Someone may remember an event better than someone else, without anybody having to

be wrong. Moreover, one may even say, of a person who remembers something false about a past event, that she remembers it better than another person who only remembers something true about that event. Suppose that Sam and Tom witness the exact same event, say, a little baby being picked up by his mother at the park when he was crying. Sam reports the event thus:

(9) I remember that the baby was crying and his mother picked him up.

All that's true. But Sam does not remember what the baby was wearing. If pressed, he wouldn't be able to tell you. Tom remembers the same details that Sam remembers, but he also remembers that the baby was wearing an overall with an illustration. Here's the rub: he remembers that the illustration depicts a boat but it actually depicts a plane. Tom reports the event thus:

(10) I remember that the baby was crying, his mother picked him up, and he was wearing an overall with a picture of a boat on it.

I take it that most people would say that Tom remembers the event better than Sam, even though Sam's report, unlike Tom's, is true. The reason, I surmise, is because we evaluate the content of our episodic memories in terms of degrees of correctness. Memories, in that sense, are like perceptions. They include a plethora of sensory details that one can get more or less right. When you remember a particular episode, you bring to mind a complex sensory scene that represents the experienced event more or less accurately.

Now, the fact that the contents of our memories can be correct or incorrect however, does not mean that they are propositional. As Tim Crane (2009) has recently argued for the case of perception, it does not follow, from the claim that a particular perceptual experience represents the world as being in a certain way either correctly or

incorrectly, that the content of a perceptual experience can be true or false. Correctness, unlike truth, admits of degrees. But propositional contents can only be true or false (McDowell, 1994). Thus, Crane argues, perceptual contents aren't propositional. I think the same argument, mutatis mutandi, applies for episodic autobiographical memories. As in the case of Sam and Tom, when we bring to mind a past event, the intentional content we are aware of is rich and complex; it represents an experienced event in ways that range from very sketchy to very faithful. Remembering just the gist of an experienced event is, indeed, different from remembering in a more detailed way, but this difference isn't captured by saying that one content is false while the other one is true.

A possible reply here is to say that this is merely an expression problem. One could say that no matter how richly detailed and complex an episodic memory may be, there is a possible sentence (or utterance) the content of which is the same as the content of the memory. If they had the right sorts of linguistic resources, then Tom and Sam would have been able to properly express the contents of their respective memories by way of two different sentences, each one of them expressing different propositions. But notice that this reply capitalizes on a loose understanding of "same content". The thesis cannot be that the content of the possible sentence is *identical to* the content of the memory. Stipulating such an identity simply begs the question. The obvious alternative is to say that the sentence perfectly *describes* the content of the memory. But the description of an intentional content is not identical with the content itself (Crane, 2009). Not only are there many ways one can describe a particular intentional content, it is also unclear whether any description can manage to capture the intentional content in its entirety without any informational loss.

This last consideration—reminiscent of Dretske's (1982) distinction between informational formats (i.e., pictures are analogue, propositions are digital)—suggests a last argument against the claim that the content of our episodic memories is propositional. When philosophers claim that a particular content is propositional, they normally imply that the person entertaining such content must be able to deploy the concepts constituting the proposition. In other words, a person cannot entertain a propositional content unless she possesses the concepts required for believing such proposition. Thus, if we can show that one could have an intentional content for which one does not have the relevant concepts, the possibility of non-propositional intentional contents follows. This is precisely what Martin tried to do with the example of Archie and the cuff link (Martin, 1992; the example is originally Dretske's, 1969). The example invites us to imagine that Archie is looking for his cuff link. He looks in the drawer but fails to notice it. He was in a hurry, as he had to get to an important dinner. On his way to the dinner he revisits his searching the room and he sees himself, as it were, looking into the drawer. Having a good visual memory, he conjures up a detailed image of the drawer. Now he realizes that the cuff link was in the drawer. He had failed to notice it before because he did not attend to it.

The point of Martin's example is that both during perception and during recollection, the drawer looked to Archie in a certain way. But—Martin contends—if the content of Archie's perception of the drawer was propositional, then he would have been able to form the belief that the cuff link was in the drawer. However, he did not form that belief. Only later, upon recollection, was Archie able to form the belief that the cuff link was in the drawer. But given that the content of the recollection fully depends upon the

content of the original perception, Martin concludes that the way the drawer appeared to Archie during perception must have been the same way in which it appeared to him during recollection, even though he was only able to form the belief that the cuff link was on the drawer when he remembered it. Therefore, Martin concludes, Archie entertained a mental content during perception that wasn't constrained by the concepts he deployed during the perceptual experience. The content of his perception was non-propositional.

Unfortunately, there is an obvious problem with Martin's argument: it is simply false that one's recollection fully—and solely—depends upon the perceptual experience it is a recollection of. Psychological evidence clearly shows that our recollections are highly reliant on current conditions of recall, including one's beliefs, intentions and even prior related perceptual experiences (as I extensively argue in next two chapters). As a result, it is perfectly plausible that in remembering the drawer, Archie was actually cobbling together pieces of information that came from other sources, in addition to the perceptual information of his perceptual experience while he was looking at the drawer. Thus, Martin's claim that the intentional content during perception and during recollection is the same is simply not warranted.

Nonetheless, Martin was onto something important. In order for his example to succeed, he needs to show that there is some kind of information that is made available to the subject during recollection, due to her having experienced it during encoding, but which nonetheless was unavailable for the subject to form a belief about it at the time. And anyone familiar with the literature on memory would recognize that this is basically the definition of priming. In a classic study, Tulving and collaborators (1982) presented participants with a list of words during a study session. Later, at test, participant saw

word fragments that could have been completed by more than one word. Relative to controls, participants who were exposed to the list of words during test tended to complete the word fragments with words from the studied list even though they had no conscious recollection of having seen those words before. This study—as well as its innumerable subsequent replications—suggests that forming the conscious belief of having seen a particular stimulus during encoding is not necessary for that stimulus to have an effect later on during retrieval. In fact, there is no need to even form a conscious belief about the stimulus during encoding for it to have a subsequent effect during retrieval. Following Tulving et al's lead, Forster and Davis (1984) presented participants with words subliminally—i.e., words that were rapidly flashed (~50 milliseconds) and immediately masked by different words with a longer exposure time (~500)—in order to determine whether conscious awareness during encoding was required for priming. The results showed that words subliminally presented also generate priming, so neither during encoding nor during retrieval is conscious awareness required for encoded stimulus to produce a differentiable effect during recollection. Incidentally, these studies, as well as Martin's example, highlight the role of attentional allocation during recollection. What we are consciously aware of seems to be determined not only by what we pay attention to during encoding, but also what we focus on during retrieval. Since Archie did not attend to the relevant location in the drawer when encoding, he was unable to form a conscious belief about the presence of the cuff link at the time. It was only when he focused on the right region of his memorial content that, upon retrieval, he was able to form the belief that the cuff link was effectively in the drawer. Likewise, a case can be made to the effect that when we fail to attend to a particular stimulus, it does not reach conscious awareness

via working memory, in which case the information isn't stored in long-term memory for subsequent explicit retrieval. Attentional allocation thus becomes a necessary condition for encoding material in long-term memory (This point will become relevant below, and it will be extensively discussed in the next chapter.)

Let's take stock one last time. I began this section pointing out that the arguments put forth in the second section only affect an ontologically loaded reading of the relational view. An ontologically innocent semantic reading of such view can still hold that remembering is a propositional attitude on the basis that memorial contents are propositional. So I suggested that there is still reason to believe that remembering is not a propositional attitude from the fact that memorial contents may not be propositional. I offered two arguments in favor of that claim, both aimed at showing that memorial contents aren't truth-bearing. The first argument targeted the factivity constraint, the acceptance of which could support the claim that memorial contents are truth-bearing. The second argument tried to show that memorial contents admit of degrees of correctness, as opposed to the all-or-nothing view of truth-values assumed by propositionalists. I ended up this argument suggesting that, by not being propositional, memory contents may include non-conceptual information.

4. Conclusion: The challenges

In this chapter I tried to show that the view according to which remembering is a matter of being in a relation with a proposition is wrong. In the first part, I argued against the relational aspect of this view by pointing out that relationalism faces one of two problems: either it has a hard time accounting for the way in which propositions can play

a causal role in psychological explanations (the problem of traction) or it has a hard time finding empirical evidence in its support (the problem of evidence). In the second part, I argued against the claim that the content of our memories is representational. In doing so, I suggested that memorial contents may not be truth-bearers, that they can be more or less accurate, and that attention seems to play a critical role in making perceptual information available for episodic encoding and retrieval.

Each one of these claims, however, constitutes a challenge, for denying that remembering is a propositional attitude comes with a price. As a result, a successful account of remembering must give us the tools to explain how memories can cause behaviors and other psychological states in virtue of their content. Additionally, to be empirically sound, it must make sense of the different degrees of accuracy, and it should explain the role that attention plays in making some perceptual information available for encoding and retrieval. Finally, a successful account of remembering must also be able to make sense of another phenomena the propositional attitude account managed to account for: the fact that mental contents are sharable. Two or more people can believe the same belief, wish the same wish and, presumably, remember the same memory. I confront these challenges in the next chapter.

4. Memory, Attention, and Joint Reminiscing

If we consider evidence rather than presupposition, remembering appears to be far more decisively an affair of construction rather than one of mere reproduction.

F.C. Bartlett (Remembering, 1932: 205)

1. Introduction

I went to a primary school far away from home. As a result, I often had to endure very slow bus rides of over one hour in heavy traffic. To avoid boredom, the other kids and I used to play a game called "Veo Veo" ("I See, I See"). It was rather simple. One of us, gazing through the window of the bus, would glance over the busy scenery of the city. Meanwhile, everybody else would keep their eyes closed. Eventually, the kid who was surveying the scene would mentally single out a particular object and he would say "Veo Veo". That was the sign for the rest of us to open our eyes and ask "¿Qué ves?"—"What do you see?" He would then give us one clue, a particular aspect of the object of his attention, and we would try to guess the object he had in mind. We could ask up to five questions of the form "Does it have an X?", where X was a property of the object we thought the kid was attending to. If the kid said "no", that meant we were focused on the wrong object, so we would have to attend to a different one with only four questions remaining. If the kid said "yes", then one could either keep asking—just to make sure one had the right object in mind—or one could try to guess what the intended object was. If

you were wrong, you'd have lost. But if you guessed correctly, you'd get to pick the next objet. The point of the game was then to be the first one in attending to the same object as the kid who got to pick it.

What we were doing was an exquisite exercise in what psychologists call *joint attention*: our capacity to attend to the same object while also realizing that the other person is attending too (Moore & Dunham, 1995). Consider the moment in which the kid who had mentally selected the object in his visual field realizes that another kid guessed correctly. How does the former know that his thought refers to the same object the latter has in mind? First, both of them need to have the selected object in their visual fields. This, obviously, is not enough. After all, the other kids, at some point or another, had the selected object in their visual fields. Additionally, they both needed to single that object out of its surroundings; both of them must have selectively attended to it. But, once again, this isn't enough. Another kid, whether playing or not, may have been attending to that very object, at that precise moment, without realizing that the object of his or her attention was the object chosen by the kid who was picking it out. What is required, then, is a kind of attentive triangulation, whereby both kids are aware of each other and of the object, plus the recognition that each other knows that the object they are attending to is the chosen one. According to Campbell (2002, Ch. 8), this attentive coordination makes the other subject, as well as the object, a *constituent* of the content of their joint mental state.

But now suppose that we want to play a different game, one that we may call "I Remember, I Remember". It is just like "I see, I see" except that, in this version, one of the participants remembers a particular object or event and then the others have to guess

what he has in mind. It sounds much harder, doesn't it? After all, unlike the case of perceptual joint attention, the alleged constitutive relation between the perceivers and the object of attention cannot be met. In the case of joint reminiscing, the intentional object is not present—it may not even exist. Moreover, unlike perceptual joint attention, it isn't required that both subjects were ever at the same time in direct contact with the object of their memories. For instance, one can jointly reminisce about an old professor with another alumnus of the same school that one just met. Of course, it may be possible that, in the course of jointly reminiscing, both realize that they shared a class, but this does not mean that, at the time, they both were aware of each other jointly attending to the professor.

Surprisingly, though, we engage in joint reminiscing all the time. What does it take for us to engage in joint reminiscing? How can two or more people jointly entertain the same memorial content? Moreover, how can we jointly refer to an object or an event that is long gone? One obvious possibility—if you are a philosopher of mind, that is—is to say that both rememberers relate to a mental sentence referring to the same proposition, and that both relate to it with the same (memorial) attitude. For reasons I discussed in the previous chapter, I am skeptical of this response. In this chapter I want to offer a different explanation. Departing from an idea first suggested by Russell (1913), and later on developed by Campbell (2002), I suggest that our capacity to refer to particular objects or events during memory retrieval crucially depends on our capacity to direct our attention inwardly toward the relevant components of our memory experience. Inward attention is not enough, however. In addition, I argue that we must possess the ability to refer to a certain object by way of pointing to a different one. Finally, I claim

that in order to jointly reminisce, we must have the capacity to direct someone else's attention inwardly toward the relevant aspect of the mental representation we want them to focus on, so that they get to know which past object or event we are deferedly ostending. I explain each element of my account in turn.

2. Memory and mental ostension

Let me start with a methodological digression. One of the morals I wanted to draw from the previous chapter is that we should distrust the strategy of trying to draw conclusions about the nature of our psychological states on the basis of semantic analyses of the expressions with which we report or ascribe them. Language is the social product of an evolved psychological capacity, and I know of no good reason to believe that it manifests equally in all cultures, that it has a universally shared logical structure, or that such structure reflects reality. Instead of this *language-to-mind* approach, I want to follow a different strategy. Suppose we want to understand what is the mental state referred to by a memory report such as:

(1) I remember that you were wearing a hat at my party.

Instead of thinking how the world should be for (1) to be true, I suggest that we should try to figure out, on the basis of what we know about the world, how we manage to truthfully express mental states with sentences like (1). This *mind-to-language* approach involves at least three stages: (a) understanding the content of a memory experience, (b) understanding how we manage to truthfully express such content, and (c) accounting for the relation between the content of a memory experience and its object. The purpose of this section is to give an account of (a) and (b). I leave (c) for the next section.

Let us begin by putting (1) in the context of a memory experience. I am grocery shopping, strolling down the aisles, when all of the sudden I hear a female voice, behind my back, calling my name. Think of what happens as a result of my hearing this brief sequence of phonemes. First, since I was silently focused on a particular visual scene, the noise made my attention shift from the shelves onto the stimulus behind my back. Given the silence around me, any auditory stimulus would have done that, of course. But this noise was a particularly relevant sequence of phonemes: it was my name. Had it been any other sound against a noisy background, I may not have heard it. But my brain is attuned to certain noises that are socially relevant for me, like my name, and as a result it makes me conscious of them even if I had been sensitized to background noises of equivalent pitch and volume (Triesman, 1983). This shift of attention to an exogenous stimulus, which my brain had already recognized as socially relevant, in turn shifts the cognitive mode I am in to what psychologist Endel Tulving called "retrieval mode": a mental state in which I am poised to retrieve information from memory. This occurs as I turn my back toward the source of the stimulus, no more than 400 or 500 milliseconds after its onset. The face, the tone of voice and the mannerisms of the woman behind me constitute the perceptual cues with which I now try to recognize her—but to no avail. Perhaps noticing my (micro) facial expression of confusion, she appends her call out with a new utterance: "We met last week at your party". This new string of auditory information, added to the already in-process perceptual cues, reactivated particular memory traces toward which I now turn my attention inwardly. Now I am covertly surveying bits and pieces of the different visual scenes I am being conscious of. These images are presented to me as blurry snapshots, maybe even quick footages, of scenes featuring my house and my

friends in situations I recognize as having happened last week during my party. All of the sudden, there is a match between the perceptual cues and the memory trace I'm aware of—a phenomenon Semon (1904/1921) called "ecphory". My attention has been focused upon a particular region of a scene in which I see a person that highly resembles my interlocutor; she's bending down, picking something she seem to have dropped on the floor, and she looks at me. I see she's wearing a hat, and I utter (1). The woman in front of me smiles approvingly. No more than four seconds elapsed since she said my name.

The first thing to notice is that the information I am aware of when I finally remember that she was wearing a hat at my party presents to me in a way that can be accurate or inaccurate; I may not remember the color of the hat, for instance, or she may not have been picking something up. In this sense, then, my memory experience has intentional content. Now, how did I become aware of that particular content? The answer to that question actually involves two parts. First, we need to understand how the brain manages to retrieve this particular memorial content, and then we need to account for how the retrieved content becomes conscious. Let us start with retrieval. Most philosophical accounts of memory retrieval have been mere speculations based on the commonsensical idea that experiences are somehow saved in a metaphorical storehouse, where they lose vivacity over time as though they were accumulating dust, awaiting their eventual retrieval during recollection. Recent developments in cognitive psychology and neuroscience have shown that this view is very much mistaken. For one, memory consolidation—i.e. the physical process by means of which the brain changes so as to encode experienced information in a memory trace—is a highly selective process. Not all the information that was initially perceived is encoded, and not all the information that is

encoded is available for retrieval. Much of our sensory information is lost due to inattention and working memory limits, as well as normal decay caused by lack of rehearsal and, apparently, selective consolidation during sleep (Paller & Voss, 2004). In addition, the encoded information does not remain stable over time. Almost four decades of research in the cognitive psychology of false and distorted memories have shown that, during retrieval, memories become malleable and prone to being contaminated by extraneous information (Roediger, 1996). Finally, evidence also suggests that events that were only sketchily encoded can nonetheless be remembered with detail via pattern-completion processes that fill in the missing information in surprisingly reliable ways (see McClelland et al., 1995, and next chapter).

As a result, there is now wide consensus among neuroscientist regarding the *reconstructive* character of our memories (Schacter et al., 1998; Schacter & Addis, 2007). Remembering does not consist of the exact reproduction of previous experiences, but rather of the reconstruction of previously entertained mental contents by way of reactivating the brain regions that processed them during encoding (Rugg, 2009). Reactivation, of course, is not all there is to it, as we need to tell apart memories of previous events from experiences of current events. The brain manages to do that by way of incorporating, during retrieval, brain regions that were not involved during encoding, and also by redeploying some of the same regions for different purposes. In particular, whereas encoding recruits the sensory cortices and the medial temporal lobes, retrieval additionally recruits pre-frontal and parietal cortices.

To better understand what I mean by memory contents being reconstructed during retrieval, let's go back to the previous example to see how my memory of the woman

with a hat gets first encoded and then retrieved. Suppose that, at my party, I did in fact meet the woman with the hat, and I did in fact attend to her picking something up from the floor. My sensory cortices first processed this fleeting perception in a distributed manner (i.e., visual information in occipital cortex, auditory information in auditory cortex, etc.). Since I did pay attention to her, and to the fact that she was wearing a hat, this perceptual information made it into my working memory, and in turn it was bound together by the hippocampus as a single, unified event.[15] Neurophysiological evidence suggests that the area CA3 of the hippocampus carries out this binding, storing a sort of index of the episode (McClelland et al., 1995). However, this index does not include any sensory information per se. Instead, it records the manner in which the pattern of sensory activation during my perceptual experience occurred in order to reenact it at retrieval. Thus, when presented with a cue—in this case, the utterance of my name—the brain gets into retrieval mode, which apparently is subserved by the fronto-polar cortex (Rugg & Wilding, 2000). Using every piece of sensory data as a potential cue for retrieval (e.g., the woman's voice, her physique, etc.) my brain tries to get the hippocampal index to reactivate a perceptual pattern. Yet it is only upon the pronunciation of the right cue—in the example, a contextual-semantic piece of information—that ecphory is achieved, and the right index gets to reactivate, more or less, the pattern of neural activity in which it was when it first perceived the woman at my party. Incidentally, the fact that every time a memory trace is reactivated it occurs in a different neuronal and experiential context (e.g., the mental state one is in at the time of recollecting), means that each reactivation of a memory is also an instance of reconsolidation (Moscovitch et al. 2005; Hardt et al.,

[15] Had I not paid attention to her, sensory information would not have made it to working memory, so it would not have been consolidated in long term memory. Aspects of the event could, nonetheless, be retrieved subsequently, but only in a non-declarative manner

2010). This helps explain why retrieval makes memories vulnerable to distortion. In sum, the content of my memory is the result of a complex process of sensory reactivation in which sub-personal level representations are bound together in order to reconstruct the perceived content during retrieval.

Now, how did I become aware of this content? More specifically, how is it that the retrieved content presents to me as being about this woman wearing a hat at my party? My suggestion is that it becomes conscious when I covertly focus my attention on the region of my retrieved representation depicting the woman's hat as she was turning toward me. Only then was I able to (sub-personally) match the retrieved content with my present perception, and only then was I able to recognize her as the woman I am talking to right now. Additionally, it was only when the attended content of my representation became the focus of my conscious experience that I was able to say that I remember that she was wearing a hat at my party. In other words: it was by way of mentally delineating a particular region of my intentional content that this aspect of the scene was experientially highlighted to me, and it was this highlighting that made it available to my conscious reporting. This is basically the memorial equivalent of what Campbell (2002) calls the *Causal Hypothesis* for visual perception: "When, on the basis of vision, you answer the question, 'Is that thing F?', what causes the selection of the relevant information to control your verbal response is your conscious attention to the thing referred to" (p.13). My claim is that the same mechanisms by means of which you consciously attend to a region of space are responsible for the experiential highlighting in a memory experience. I call this experiential highlighting "mental ostension" (tantamount to what Prinz, 2007, calls "mental pointing"). To mentally ostend or point toward an

aspect or a feature of a mental content is to focus one's attention inwardly toward such aspect. And mental ostension is the mechanism by means of which the mental content—or the region of the mental content—we attend to becomes available for conscious report.[16]

This hypothesis finds strong support in evidence coming from cognitive psychology and neuroscience. Behavioral studies show that attentional mechanisms gate information into working memory. Since having information in working memory is a condition of possibility for its being verbally reported, failure to attend to a particular aspect of the content would result in failure to report such information verbally. Studies using a dual-task paradigm during retrieval show a significant reduction of recollection in divided attention relative to full attention (Fernandes & Moscovitch, 2000). This effect is even larger when the secondary task taps at the same kind of material as the primary task. So, for example, if the retrieval task is verbal, a word-based secondary task would be more detrimental to successful recollection than a digit-based or a picture-based task (Fernandes et al. 2005). Further evidence comes from neuropsychological studies. Damage in posterior parietal cortex usually causes attentional deficits, with hemispatial neglect being the most typical one. Patients with neglect fail to attend to the hemispace contralateral to the lesion even though they show no sign of perceptual deficiencies. When asked to describe a room or an object, for instance, they basically ignore everything on the side opposite to the locus of the lesion. Interestingly, the same occurs when they are asked to *remember* a familiar place. In a classic study, Bisiach and Luzzatti (1978) asked a patient with severe hemispatial neglect to remember the main

[16] Although I have endorsed the view that attention is the mechanism for consciousness (De Brigard & Prinz, 2010), all I need to be committed to here is the weaker claim that retrieved contents are made conscious when inward attention makes them available for working memory.

square in Milan, the city he lived in all his life. Although his language capacities where impeccable, his report omitted all the buildings to the left of the square when he remembered it facing one direction. Then he was asked to imagine crossing the square and turning back, so that now he'd be facing the opposite side. Again, he failed to report the left-hand buildings, even though those were the buildings he had just reported! A final piece of evidence comes from recent neuroimaging studies, as activation in parietal cortex—strongly linked to attentional mechanisms—is one of the most frequent findings in PET and fMRI studies of episodic retrieval (Rugg & Henson, 2002).

This last piece of evidence speaks to another important component of this hypothesis: it accounts not only for voluntary but also for *involuntary* recollection. Philosophers and psychologists are mostly interested in memories that are deliberately retrieved. However, many of our memories just pop into our minds spontaneously—sometimes to a debilitating extent, as it occurs in PTSD (Berntsen, 2009). According to the view I am suggesting, involuntary memories, just like voluntary memories, become conscious when their contents are made salient by focusing our attention upon them. But there are two sides to saliency: a certain region of our memorial content can become salient because we pay attention to it, but we can also pay attention to it because it is salient. This two-way directionality of saliency has been explained by cognitive scientists in terms of two distinct attentional mechanisms: top-down and bottom-up. The former refers to a deliberate and controlled attentional search, and it is thought to be carried out by a dorsal fronto-parietal network. Conversely, bottom-up attention refers to unexpected and spontaneous detection of stimuli, and it is supported by a ventral fronto-parietal network (Corbetta and Shulman, 2002). Recent neuroimaging and neuropsychological

studies have shown that both attentional mechanisms are differentially engaged during retrieval of voluntary and involuntary recollection (Cabeza, 2008). Consistent with the attentional dissociation, voluntary recollection recruits the dorsal fronto-parietal network to a greater extent than involuntary recollections which, in turn, appear to engage the ventral fronto-partietal system. Neuropsychological data also supports this view. A recent study of two patients with damage to ventral regions of their parietal cortices showed severely diminished free-recall of autobiographical memories relative to controls, but normal recall when explicitly cued (Berryhill et al., 2007). This suggests that goal-oriented direct attention drives voluntary recollection, but free-recall—and presumably involuntary recollection—engage bottom-up attentional mechanisms. Thus, during voluntary recollection mental ostension is the result of top-down attentional mechanisms, whereas in involuntary recollection, it is achieved via bottom-up attentional capture.

In sum, when I retrieve episodic information I reconstruct sub-personal level representations by way of binding them together, a process that involves an interaction among the parietal cortex, the medial temporal lobes and the pre-frontal cortex. The content that becomes available for conscious report is that which I have directed my attention toward. Voluntary goal-directed memory retrieval is subserved by top-down attentional mechanisms, whereas involuntary spontaneous recollection is driven by bottom-up attention. Finally, when attended contents are consciously experienced they become available for working memory, which in turn allows for the phonoarticularoy system to engage in vocal or subvocal linguistic production. When the resultant verbal production is vocal it constitutes an utterance which—modulo communicative intentions and good faith—aims at reporting the mental state of which it is an effect. This is,

therefore, a rough sketch of what happens when I utter (1) in a situation like the exemplified above. And now that we have an rough idea of what the content of a memory experience is (point (a) above), and how we manage to truthfully report it (b), it is time to tackle question (c) regarding the relation between the intentional content and the object of my memory experience.

3. Remembering as deferred mental ostension

If the above account is on the right track, remembering is a matter of mentally pointing toward an experiential content, and it is by way of mentally pointing toward such content that we can make it available for conscious reporting. As it stands, however, this view poses a difficult question. In the case of perception, the object that is mentally ostended—or 'experientially highlighted', in Campbell's terms (2002)—is in direct contact with the perceiver. Indeed, in the relational (realist) view that Campbell puts forth, the object becomes a *constituent* of the experiential content. Thus, for Campbell, there is no need to separate the intentional content and its object when it comes to making them available for consciousness. However, in the case of memory, the object of one's recollection isn't in direct contact with the rememberer. In fact, the object of one's memory not only is not present when we remember it, it usually no longer exists. How can we be aware of an object or event with which we are no longer in direct contact?

One possibility is to go the *direct realist* route (Reid 1785/1849). According to this view, remembering is tantamount to direct perception, in that the intentional objects are directly apprehended. Intentional contents, particularly representational contents, are thus disposed of. For direct realism has it that remembering is just like perceiving, except

that its objects—i.e., that which is remembered—do not exist in the present: they exist in the past. Although relatively popular among some philosophers (Laird, 1920), and even some psychologists (Gibson, 1979), direct realism for memory (although not for perception) fell in disrepute. It faces, after all, difficult obstacles. For one, direct realism suggests an analogy between memory and perception but it does not specify the extent to which they are similar, or how to accommodate their obvious differences. Memories, for instance, are phenomenologically different from perceptions, and they are usually coarser. Other functional dissimilarities go beyond mere phenomenology. Memory and perception also differ in the capacity to provide us with discriminatory information. For instance, while we can visually tell apart very similar shades of red when perceived simultaneously, we are at random if we are to rely solely on memory (Halsey & Chapanis, 1951). In addition, memories decay and are often blurry and lifeless. Also, we tend to remember fewer details of non-traumatic old memories than of traumatic or recent ones, even though we perceive both kinds of events equally well. Non-salient events tend to be more easily forgotten than salient ones, even if the salient ones occurred much before in time. It is hard to see what the equivalent of this kind of saliency effect would be for perception. Finally, there is the problem of false memories. Empirical evidence shows that the mechanisms of veridical and non-veridical remembering are very much the same, and that many of our veridical memories are actually the result of the same mechanisms that give us non-veridical memories (see Schacter, 1995, and chapter 4). But non-veridical memories are about events that never occurred. As a result, the direct realist would have to explain not only how can memory be in direct contact with an event that no longer exists—or that exists in the past (whatever that means)—but also with events

that never existed. Some metaphysical maneuvering could potentially solve these issues, but it is unclear whether this is a price we want to pay, the alternative being to simply accept the existence of representational contents (Furlong, 1948).

Aware of these problems, Campbell suggests a different non-representational alternative, based upon McCormack and Hoerl's notion of *temporal decentring*: "The ability to temporally decenter is the ability to consider alternative temporal perspectives on events and to understand the relationship of these perspectives to one's current perspective" (McCormack & Hoerl, 1999; see also Evans, 1982). Accordingly, Campbell suggests that our capacity to refer to remembered objects or events depends upon our capacity for temporally decentering. It is only when we acquire the capacity to temporally decenter that we can grasp the truth-conditions of judgments tensed at times different from when they are uttered. According to this view, in order to understand the sentence:

(2) I see that you are wearing a hat at my party

when uttered in the presence of the object of attention (i.e., the person wearing a hat at my party), we only need to be able to grasp the truth-conditions of the judgment as it applies to the current situation. But in order to understand (1) we need to be able to move away from the current temporal situation, and grasp the truth-conditions of the judgment *as if* it had been made at a different time, namely the relevant moment in the past. Therefore, there is no need for direct contact with the past object, nor a reference to any intermediary mental representations. All that is required is the acquisition of a particular skill—i.e., temporal decentering—so that we can refer to the object of our conscious recollective experience as if we have been talking about it at a different time (Campbell, 2002: 181)

Although I am not completely unsympathetic to this view, I find it unsatisfactory for two reasons. First, according to McCormack and Hoerl (1999), the development of episodic memory depends upon our acquisition of temporal decentering, which in turn depends upon the acquisition of the concept of personal/perspectival time. Although this hypothesis seems to fit much of the data in the developmental literature, it has a hard time accommodating data coming from neuropsychology. After all, individuals with amnesia are perfectly capable of using personal/perspectival concepts, and thus are perfectly capable of temporal decentering, even though their episodic memory is damaged (Rosenbaum et al., 2006; Craver, in preparation). This suggests that temporal decentering is not required for episodic memory. Second, defining something as a cognitive skill or capacity does not preclude it from requiring representations, whether conscious or unconscious. In fact, even motor skills appear to require proprioceptive representations in the somatosensory cortex, the neurons of which code for specific postural situations. Explaining the specific mechanisms of a particular skill may still require the postulation of intermediary representations.

This, I think, is precisely what happens when we refer to the objects of our memories. I believe that we can keep the intuition that mental ostension is the mechanism by means of which we can refer to the object of our memories without having to accept temporal decentering. Instead, I suggest that what allows us to refer to past objects or events when we are consciously attending to a particular mental content that presents itself as being about a previous experience, is the covert equivalent of our overt capacity to ostend or demonstrate deferredly. Notice that the root of the problem we are facing is that what we mentally point at when we remember is not identical to what we refer to.

Consider (1) again. When I utter (1) I am not talking about my mental experience but about the event in which this woman was wearing a hat at my party. It is an event that no longer exists. But what I am inwardly attending to—what I am mentally pointing at—is a region of the intentional content I am being aware of, right now, as I am having the mental experience of remembering the woman wearing a hat at my party. Schematically, if '*p*' stands for the intentional object of my memory, and '*r*' stands for the intentional content of my memory, according to my proposed account, when I remember that *p* I am talking about *p* while pointing at *r*.

Linguistically, the phenomenon of ostending at a certain thing '*r*' in order to refer to a different thing '*p*' is known as *deferred ostension* (Quine, 1968: 194). Consider the classical example due to Evans (1981: 199). We are walking down the street and I point toward a parked car covered with parking tickets. Pointing at it I say 'That man is going to be sorry'. The intuition here is that even though I am pointing at the car—that is, even though my demonstration (Kaplan, 1989) is directed toward the car—the object demonstrated or referred to is not the car, but the *owner* of the car. Or consider the situation in which I point to a set of footprints and say 'He must be giant!', or the case in which I am holding a copy of *The Confederacy of Dunces* and say 'he's my favorite author' (Borg, 2002). These too are cases in which I am pointing at something (e.g., footprints, a book) while refereeing to something else (e.g., whatever animal left the footprints, John Kennedy Toole).

My suggestion is that the same sort of phenomenon occurs when we remember episodic memories. To understand what the objects of our memories are, and consequently to be able to talk about the objects of our memories, we first learn how to

mentally refer to something that is not perceptually present in one's environment but that nonetheless is present in our conscious experience. Developmental psychologists have debated for decades whether preverbal children have episodic memory. Some developmental psychologists suggest that visual pair comparison tasks, whereby babies are presented with novel versus familiar objects and their kicking rates are measured, are good indications of the origins of episodic memory. However, many others disagree, as it is always possible to interpret this paradigm as tapping at implicit rather than explicit memory. Nonetheless, probably all developmental psychologists agree that deferred imitation of action sequences does in fact demonstrate the emergence of episodic memory (Barr et al., 2005). Moreover, older adults with medial temporal lobe damage as well as individuals with developmental amnesia have trouble with this task, further suggesting its intimate relation with episodic memory (McDonough et al., 1995; Adlam et al., 2005). In this paradigm, infants are shown a relatively unusual sequence of actions with a particular object. For instance, the experimenter may show the child that in order to get the key out of the box she needs to first hit the box three times with the tip of the magic wand and then once with the bottom. Then the child is either left alone (and recorded) or the experimenter leaves for a few minutes and comes back with the wand, asking the infant whether she can get the key out of the box. Prior to 6 months of age, infants are completely incapable of reproducing previously learned action sequences. There is some evidence that they can perform deferred imitations of brief sequences after 6 or 7 months of age, as long as the retention interval—i.e, the elapsed time between study and test—is kept fairly short (Barr et al, 1996). Gradually, children learn how to perform action sequences that are increasingly more complex, that have longer retention intervals, and

that are retrieved with less specific cues (Hayne et al., 2000). By the second year of age, deferred action sequences are pretty much established.

Notice that, prior to 6 months of age, infants are capable of pointing. If one shows a 4 month old the magic wand, she can point at it. Nonetheless, she does not see it as related to anything that happened before. It is just another object in the visual field, however interesting it may be. After 6 or 7 months of age, though, the infant appears to be able to see the magic wand as something more than a mere present object. She sees the magic wand as related to a previous event. It is no longer an isolated visual stimulus: the wand becomes a cue. It becomes the sort of object the perception of which can elicit the mental content that presents to the infant as this or that sequence of previous actions. Now, the experimenter is able to ostend at the wand, while the infant perceives it, and ask for the right sequence of actions: "Can you get the key out of the box?" The fact that the infant can indeed come up with the right sequence of actions strongly suggest that she knows that one can talk about a previous 'p'—a sequence of actions—while pointing at a present 'r'—the magic wand.

As time goes by, the perceptual cues can become less and less concrete, that is, less and less similar to the perception of the original event. Eventually a pretend wand can elicit the memory, then just the waging of a finger, the uttering of a word. Suddenly, the demonstration of the cue, and perhaps the cue itself, becomes irrelevant. All that matters is that it can elicit the retrieval of the right sort of mental content, and that it can experientially highlight the relevant property of the resultant conscious experience. Neither my hearing the woman's voice nor my seeing her face succeeded in triggering the right memorial content. It was only when she gave me the contextual information that

ecphory occurred, and the right intentional content was then retrieved[17]. Now, the sensory information I have been presented with mentally highlights a certain aspect of that content, which is experienced by me as a reinstatement of the perceptual event of seeing this woman bending down, and turning her face at me, wearing a hat. Mental ostension is, thus, an acquired skill, and deferred mental ostension is a way we learn to use mental pointing to refer to something else—usually that which caused the retrieved intentional content to begin with[18]. My suggestion, therefore, boils down to no more than this: we can talk about the intentional objects of our memories because we can refer deferredly to them by mentally pointing toward the intentional contents we experience when retrieved by the right cue. Remembering a past event is a case of deferred mental ostension.

4. Joint reminiscing as concerted mental deferred ostension

Let me review what I have said so far. In the first section I argued that memorial contents are reconstructed out of sub-personal level representations via a process of

[17] Presumably, a working hippocampus is required for ecphory to take place. Absent the right sort of hippocampal index, the process of pattern completion required for sensory reactivation is hindered, so no mental content upon which to turn one's inward attention is retrieved. This would explain why individuals with medial temporal lobe damage fail at the deferred action sequences paradigm and, incidentally, partly explains why they fail to retrieve unconsolidated memories.

[18] So far, I have only talked about mental deferred ostension, in analogy with linguistic deffered ostension. I wonder, however, if what I have said here may have also some application to the lingustic phenomenon as well. In a recent comprehensive study on deferred ostension, Emma Borg (2002) shows that the strategy of treating deferred uses of demonstratives as a different semantic kind of indexicals, is wrong-headed. She presents persuasive arguments to the effect that the differences ought to be accommodated at the pragmatic level. Indeed, she suggest that the same pragmatic rule that works for perceptual uses of demonstratives also works for deferred uses, as long as the child learns that there are more than one way to demonstrate an object. Her proposal, then, "is simply that there are lots of ways to draw an object to attention to facilitate the use of a referring expression, and pointing directly to the object is just one way amongst others—other ways which include pointing at a related object" (Borg, 2002: 509). The development of episodic memory may provide the psychological basis for one of these forms, and it is likely that other forms are similarly developed (see, for instance, Hoerl & McCormack, 2004, where learning to refer to distal causes via pointing at current perceptual events is explored).

pattern-completion that reactivates, more or less[19], the sensory cortices that were engaged during the perception of the remembered event. Then I argued that we become conscious of these mental contents when we direct our attention inwardly to them. I called that process "mental ostension" (or "mental pointing"). I suggested that the memorial contents we mentally point at are thereby reportable, as they have been poised for verbal control in working memory. Then, in the second section, I claimed that mental pointing was not enough to explain how we get to talk about events or objects that aren't in the surroundings of the rememberer. I therefore argued that the capacity to talk about something not present while pointing at something present was required for the rememberer to be able to talk about remembered events or objects. Following the convention in linguistics, I called that capacity "deferred mental ostension". Now, in this last section, I suggest that in order to acquire the capacity to jointly reminisce we need to learn how to orient our attention inwardly alongside other co-reminiscers in order to mentally ostend at memorial representations with the same contents, which in turn allows us to speak about the objects that those contents represent.

Consider, once again, the situation in which I find myself uttering (1). Imagine that, after my brief encounter at the supermarket, I ran into a friend whom I know was at my party, and the following dialogue takes place: 'I just ran into the woman who spilled wine on my carpet at the party'. 'Which one?'—my friend asks. 'I don't remember her name'—I reply. 'You mean the woman who was wearing a hat?' 'That one!'—I say. What just happened? Think of what occurred during this brief exchange. I ran into my friend and by mentioning my recent encounter at the supermarket, I shift his cognitive

[19] I say "more or less" because neurological evidence suggests that even though there is reactivation at the systems level, there are changes that occur at the local neural level. The precise relationship between the sensory reactivation at the systems level and the neural changes at the local level is unknown.

mode toward that of reminiscing. When he asks "which one?" I assume he's trying to single out a particular individual from his own memory experience of the party. In other words: my opening sentence serves as a verbal trigger for *his own* memory trace of the party. Now he's surveying, via his own top-down attentional mechanisms, his own intentional content. But, of course, he does not know whom am I talking about yet. There were many women at my party. So he asks for a distinctive feature that may help him single her out: her name. Since I don't know her name he tries a new one: the hat. He is now mentally pointing toward the region of his intentional content depicting the woman in a hat, so he asks for confirmation. My saying "that one" is as good evidence that we are talking about the same remembered object, as my saying "that one" is when we are jointly attending at the same object in our visual field.

Let me stress this point. Most discussions of perceptual joint attention appear to make the act of pointing—what Kaplan called "demonstration" (Kaplan, 1989)—essential for the process to effectively take place. But one can engage in joint attention without any demonstration overtly taking place by any of the attendees. An object can demonstrate itself, as it were, by making itself salient in one's perceptual field. Suppose you are watching a soccer game when, all of the sudden, an enthusiastic fan runs from one side of the court to the other wearing no clothes. The event did not disturb the development of the game, but it was enough to grab the attention of many people in the audience, including yours and your friend's. "Do you see *that*?" your friend asks. There is no need for him to overtly point toward the enthusiastic fan. Your attention, just like your friend's, has been disengaged from what it was focused on before—the player with the ball, presumably—and it has moved onto a new target: the zealous fan. The

demonstrated object is its own demonstration. Likewise, one can guide someone else's attention toward a particular target without having to use overt pointing. One can help the other person navigate the perceptual field using intermediate salient targets as reference points. Suppose you fail to notice the naked fanatic because it failed to disengage your attention from the soccer ball. Thus, when your friend asks you whether you've seen *that*, you rightly ask, "what?" Given the distance between the naked fan and your seats, pointing is useless. And given the fact that he's holding a hot dog with one hand, and a beer with the other, handwaving is out of the question. So he finds a landmark, a salient reference point, and orients your attention from there. "See the side referee? Draw an imaginary line from him to the goal, and you'll see what I'm talking about". Your attention has been reoriented, and now you are both jointly attending at the same target.

I believe that an equivalent process goes on in the case of joint reminiscing. I can expect my friend, whom I know was at the party, to have encoded much of the same information I encoded then. The information we both encoded isn't identical, of course. Even if we both were looking at the woman from the same side of the room at the exact same time, we both occupy different spatial locations, so our perspectives are going to differ. But these differences need not matter. Memorial contents represent their objects with varying degrees of correctness (see chapter 2), and just as there may be differences *within* a subject between the way an event was originally perceived and the way it presents itself during recollection, there may be also subtle differences *between* subjects that still allow us to talk about the same memorial content being entertained by two or more joint reminiscers. Just as in the case of the fan in the soccer game, I can guide my friend's attention to highlight a particular object or aspect of his intentional content—the

woman with a hat—so it becomes the target of his mental ostension. I can use—as in my imaginary example above—a reference to a salient feature of the object itself: the hat. But I could have also oriented my friend's inward attention using other reference points: she was sitting over to one side, she had a lovely smile, she was the last one to leave, etc. Of course, the process can also go the other way around. Upon remembering this lady, my friend may be able to re-orient my attention toward a different aspect of the woman I did not remember at the time—her perfume, say, or the fact that she had brought a delicious bottle of wine. The capacity to mutually coordinate each other's attention so as to consciously highlight the (pretty much) same intentional contents I call "concerting". Consequently, our capacity to talk about the very objects represented by the intentional contents we are conscious of during joint reminiscing would be "concerted mental deferred ostension".

Importantly, memory allows a temporal dimension of concerted mental deferred ostension that isn't present in perception: we can direct each other's attention along a temporal line. In other words, we can mutually direct each other's attention toward memorial contents depicting events that occurred before or after a certain target event. For instance, when jointly reminiscing about the woman at the party, my friend can reorient my attention toward the beginning of the party, and mentally highlight to me the fact that she was not wearing a hat when she arrived. "She must have gotten it from someone else who was already there". In fact, he could even guide my attention backwards in time, reminding me that the party was not the first time I met that woman. "Remember, about a month ago, we had that picnic at the park…" This kind of concerted mental ostension along temporal dimensions is unique to memorial contents. In fact,

empirical evidence coming from developmental psychology strongly suggests that concerted mental ostension plays a fundamental role in children's learning how to talk about their memories. In a recent paper, Hoerl and McCormack (2004) survey a series of studies in which conversations about autobiographical memories between children and caregivers are analyzed. These narratives, the authors observe, exploit causal links between experienced events in order to guide the children backwards or forward in time. Here's an example of one such conversation, between a mother (M) and her child (C):

M: What happened to your finger.

C: I pinched it.

M: You pinched it. Oh boy, I bet that made you feel really sad.

C: Yeah… it hurts.

M: Yeah, it did hurt. A pinched finger is no fun… But who came and made you feel better?

C: Daddy!

I believe this is a clear example of concerted mental deferred ostension. The mother starts off highlighting a particular mental content for the child, and invites her to explore certain aspects of that content, like the emotion she felt when it happened. Then there is a temporal exploration via focusing her attention in a particular causal link: the transition from being in pain to getting better. Mother and child are, thus, jointly reminiscing an event that occurred later in time via consciously attending to a different memorial content representing the effect of the event depicted by the previously attended memorial content. Therefore, when jointly reminiscing, attended contents can become not only spatial but also temporal reference points.

To conclude, allow me to briefly recapitulate the main points of this chapter. I started off endorsing Campbell's perceptual attention view to the effect that mental highlighting—or mental ostension, in my terms—allows us to refer to the objects or events we are aware of. Then I suggested that remembering is a matter of *inward* attention toward retrieved memory traces. When retrieved memorial traces are mentally ostended—either via top-down (as in voluntary recollection) or bottom-up (as in involuntary recollection) attentional mechanisms—the mental content of one's memory becomes conscious. In turn, mental ostension makes the intentional content available for verbal reporting. In this sense, uttering a sentence such as (1) in order to express one's intentional content at the time of recollection is tantamount to describing one's content of experience. As such, the utterance used to express the intentional content of a memory experience has to be understood as a *description* of that content, and it need not reflect the structure of the content at all (see chapter 2 and Crane, 2009). In addition, I argued that in order for the speaker to successfully refer to the object of his or her memory, the capacity to mentally point to a present conscious experience while referring to a non-present one is required. In analogy with the linguistic phenomenon, I called this capacity "deferred mental ostension". Finally, I claimed that in order for two or more people to engage in joint reminiscing, and thus to be able to successfully refer to the same experienced past events, they are required to mutually coordinate their attention toward relevantly similar regions of their memorial contents. I called this processed "concerted mental ostension". Only when there is concerted mental ostension it is possible for two or more remembers to refer to the same past event. Joint reminiscing is, therefore, concerted deferred mental ostension.

5. Is memory for remembering?

So that imagination and memory are but one thing, which for diverse considerations hath diverse names.

Thomas Hobbes, Leviathan 1.2.

1. Introduction

On October 4th, 1992, a cargo plane from the Israeli airline *El Al* crashed into an apartment building in Amsterdam, exploded and left 43 people dead and several hundred injured and homeless. The event dominated the local news for many days. Ten months after the accident, a group of psychologists led by H.F.M Crombag distributed two questionnaires among a hundred Amsterdam residents. The first questionnaire asked residents whether they had seen the footage showing the plane crashing, and whether, based on their recollection of the video, they could estimate how much time elapsed between the plane crash and the explosion. Participants' estimations varied, but 55% of them remembered having seen the footage. Only 18% reported not remembering the video at all. A second, modified, questionnaire was distributed to another group, asking the residents questions about specific details of the accident as captured by the video—for example, the angle at which the plane hit the building, the exact way it broke apart, etc. Although there were some disagreements among their answers, 66% of respondents reported remembering the video vividly (Crombag et al., 1996). Here is the rub: there was never a video; there were a few photographs, but there was no footage, no amateur

recording, not even computerized reconstructions of the accident. Most of the people surveyed simply misremembered.

Far from being an unusual result, evidence gathered over the last three decades of research in cognitive science clearly show that people are prone to misremembering past experiences (Breinerd & Reyna, 2005). So-called false memories are a systematic and common occurrence in our ordinary lives, and they present a challenge to the traditional philosophical view of the function of memory. According to the traditional view, memory[20] is for remembering, and remembering consists of reproducing the contents of past experiences. When philosophers discuss the problem of false memories, they typically try to safeguard the traditional view following one of two strategies. The first strategy is to say that if a particular mental content appears to the subject to be a memory even though it does not correspond to an experienced event, then the subject is not really

[20] Two terminological clarifications. (1) It is an unfortunate linguistic fact of the English language that the word 'memory' is so polysemous. Consider the sentence "She has an extraordinary memory". It could mean that she has a good memory-qua-cognitive-system—she may be able to store a lot of information, for instance—or it could mean that she has a memory-qua-mental-state whose content happens to be out of the ordinary. As much as possible I will try to disambiguate these senses, but for the most part, when I talk about memory, I refer to the cognitive system. (2) Philosophers and psychologists recognize several kinds of memory. What psychologists call 'procedural' or 'non-declarative memory', for instance, roughly corresponds to what Bergson (1908) and Russell (1921) called 'habit memory', and James (1890) called 'secondary memory'. 'Declarative' or 'non-procedural memory', which psychologists operationalize as the kind of memory whose contents can be consciously declared, more or less corresponds to James' notion of 'primary memory'. Declarative memory, in turn, is usually divided in 'semantic' and 'episodic memory' (Tulving, 1983). Semantic memory refers to knowledge of facts and situations about the world that we need not have witnessed; when we recall semantic memories there is no need for mental imagery associated to the place and/or time in which the remembered event occurred. Some philosophers take semantic memory as tantamount to propositional memory—i.e., mental occurrences that can be expressed as 'remembering that p'—but this is wrong. After all, there are non-semantic memories that can be expressed propositionally, just as there are semantic memories that can be expressed with a gerund. Categories based on surface grammar just don't square with psychological classifications. Finally, episodic memory refers to memory of experienced events. It roughly corresponds to what some philosophers have called 'recollective memory', 'personal memory', 'experiential memory', or 'direct memory' (Furlong, 1948; Locke, 1971; Malcolm, 1963; Martin and Deutscher, 1966; Bernecker, 2010). Although there is some disagreement as to whether or not these terms define perfectly equivalent categories, I am going to leave that issue aside. I am going to refer to the kind of memory I'd like to discuss simply as 'memory', bearing in mind that the sort of mental experience I mean to talk about falls roughly within the psychologist's definition of episodic autobiographical memory. Examples include memories about particular events in one's childhood, this or that party I went to in college, the moment in which I received my bachelor's degree, or the exact instant in which my wife said 'I do' at our wedding.

exercising her memory but rather her imagination. From this perspective, then, all those individuals that reported having seen footage of the plane crashing were not really remembering: they were merely imagining. After all, since the traditional view holds that one can only remember what happened, false memories—that is, memories that do not correspond to what actually happened—are simply not genuine memories. Accordingly, one can still say that memory is for reproducing past experiences, as cases of false memories are not the products of memory at all. The second strategy philosophers take to safeguard their view is to say that false memories are indeed the result of exercising memory—and when false memories occur, memory itself is malfunctioning. From this perspective, all the people that claimed to have seen the video of the accident were indeed exercising their memories, but they misremembered because their memory systems malfunctioned. As a result, the view that memory is for remembering remains unchallenged, as false memories are simply the product of a faulty memory, the actual function of which is to reproduce past experiences.

In this chapter I argue that the traditional view is mistaken. I contend that memory is not for remembering. There are three sections to this chapter. In the first section I review some critical findings from cognitive science and neuroscience suggesting that false memories are both normal and pervasive. I argue that trying to reconcile these results with the traditional view would have unwanted consequences. On one hand, I argue that if we follow the first strategy, and consider false memories to be the product of imagination, not memory, we may have to say that some bona fide cases of remembering are not produced by memory. On the other hand, I contend that if we follow the second strategy—that false memories are the product of memory malfunctioning—we may have

to accept the counterintuitive claim that it is advantageous to have a memory system that normally fails. In order to avoid these pitfalls, I claim that our basic assumptions about the function of memory need to be revised. I undertake this revision in the second section of this chapter, where I examine the way in which we tend to individuate cognitive functions. This analysis leads me to defend a strategy according to which the function of a cognitive faculty is determined by its contribution to the cognitive organism. In the third section, I apply this strategy to the case of memory, and I reject the idea that seeing memory as a cognitive system for reproducing past experiences is the best way of making sense of its function. Instead, I offer a picture of memory as an integral part of a larger system that supports not only thinking of what *was* the case and what potentially *could be* the case, but also what *could have been* the case. More precisely, I claim that the function of memory is to permit the flexible recombination of perceptual components of memory traces into representations of possible past events that might or might not have occurred, in the service of constructing mental simulations of possible future events. I conclude by showing how this account can accommodate the evidence that is problematic for the traditional view, how it allows us to say that ordinary instances of false recollection are indeed produced by memory, and how it preserves the intuition that many ordinary cases of false recollection are the result of a memory system that is, in fact, functioning quite well.

2. Remembering what did not happen

The idea that the function of memory is to remember past experiences is grounded in the fact that when we ordinarily exercise our memory, the contents of the resultant

mental states appear to us as being about previous experiences. This fact has led many philosophers to the natural conclusion that the function of memory is to reproduce the contents of previous experiences. What licenses this conclusion is an argumentative strategy I call *the content-based approach*. According to this approach, determining the function of a cognitive faculty is a two-step process. First one figures out the way in which the contents of the mental states purportedly processed by the target faculty are experienced, and then one surmises that the system that produces those mental states must be there for that purpose.

The content-based approach—which essentially is a functional characterization based on the domain specificity of a cognitive system (see section 2)—is widely assumed in many discussions about the function of cognitive faculties in philosophy of mind. Unsurprisingly, memory is no exception. In particular, the content-based approach is followed by most proponents of memory representationalism. The predominant view in the philosophy of memory,[21] memory representationalism, says that when we remember, we deploy a mental representation depicting an event we experienced in the past. And, since memory representationalists are also representationalists about perception, remembering is typically understood as the reproduction or revival of previous perceptual representations. When confronted with the possibility that the representational content one is aware of when trying to remember may not correspond to previous experienced events, memory representationalists usually point toward some sort of mnemonic marker

[21] Although memory representationalism is the predominant view in philosophy of memory, it is not the only one. Its most prominent contender is direct realism. According to its most general interpretation, direct realism says that when we remember we don't deploy a mental representation of the experienced event; rather we become directly aware of the event itself. Direct realism is usually associated to Thomas Reid, but it has had some partisans since (see Locke, 1971, and Warnock, 1989). However, I think direct realism is an untenable position. As I argued in chapters 2 and 4, I believe that in its most extreme version, direct realism faces difficult metaphysical obstacles. As a result, for the purposes of this paper, I take memory representationalism as the default philosophical view.

by means of which we can distinguish reproductions of representations of previous experiences (i.e. memories) from mere reproductions of representations that do not correspond to previous experiences (i.e. imagination). Mnemonic markers are phenomenological properties specific to memorial contents (Bernecker, 2008). However, memory representationalists differ in the way they characterize mnemonic markers. Some claim that when the initial perception is reproduced by memory it is accompanied by the belief that it occurred in the past (Spinoza, 1985; Locke, 1979) or that it occurred in *one's own* past (James, 1890). Others think that the reproduced perceptual representation is accompanied by a decayed feeling of vivacity (Hume, 1978) or a strong feeling of familiarity (Mill, 1829; Russell, 1921). For the purposes at hand, however, these differences won't matter. The point is that, according to memory representationalism, remembering consists in reproducing previous perceptual representations. Accordingly, the function of memory is to assure the correct reproduction of previous perceptual representations by way of preserving the structure of the original perception (Hume calls this structure "the order and position of the ideas" [Hume, 1978, 1.5.], whereas Stout speaks of memory as re-instating past perceptions "in the order and manner of their original occurrence" [Stout, 1915]). As a result, when memory functions well—that is, when it successfully preserves the structure of the original perception—the reproduced mental representation is accompanied by some sort of mnemonic marker. It is unsurprising, then, that according to the traditional view, false memories are explained as cases in which memory fails to perform its function (Kurtzman, 1983).

Not all empiricists thought that false memories were the product of a faulty memory. Russell, in particular, thought that if, as a result of having been conferred the

kind of structure that normally elicits mnemonic markers, a certain mental representation is felt to be a memory, it is not clear that we should blame memory. After all, memory's role is to preserve the structure of previously experienced ideas, not to prevent them from being endowed with new forms—that is what imagination does. Consequently, only those ideas that preserve the structure of the original perceptions can be called genuine memories. Anything else would simply be a product of imagination, regardless of whether it is experienced as a memory or not. This is basically the way Russell tackled the famous example of George IV. According to this example, George IV misremembered having been in the battle of Waterloo because his repeating stories of having been there so often made him believe that he actually was. Russell claimed that George IV's continuously asserting that he was in the battle of Waterloo was instrumental in conferring the kind of structure that, later on, would elicit mnemonic markers when entertained. But, since this structure was conferred by imagination rather than preserved by memory, his apparent remembering would not have been a case of memory's malfunction. In fact, Russell went on to suggest that "cases of fallacious memory can probably all be dealt with in this way, i.e, they can be shown to be not cases of memory in the strict sense at all" (Russell, 1912: 116-117).

The problem is that considering genuine memories to be only those mental representations that preserve the structure of the perceptions they reproduce may force us to say that we are merely imagining at times in which we are clearly remembering. Consider some of the most common distortions in our ordinary experiences of autobiographical recollection: the *field/observer effect*, the *telescope effect*, and the *boundary extension* error. The *field/observer effect* is one of the most widely experienced

and documented memory distortions (Nigro & Neisser, 1983). Most of our memories are 'field memories', in the sense that we tend to remember the events they portray from the point of view from which we experienced them. 'Observer memories', on the other hand, refer to people's tendency to remember autobiographical events from the point of view an observer *other* than oneself would have had, had that observer been present during the remembered event. Thus, when we remember observer memories, we do so from a third person point of view; we can see ourselves in the mental picture, as it were. Nearly everyone has experienced observer memories at some point in their life. Usually, highly traumatic events tend to be remembered as observer memories; indeed, most involuntary recollections experienced by subjects with post-traumatic stress disorder are reported as observer memories (Rubin et al., 2008). Another commonly experienced distortion is the *telescope effect* (Neter & Waksberg, 1964), which refers to people's tendency to remember recent events as being more remote than they actually were, and remote events as being more recent. Unless the specific date of a particular event is included in the content of the memory so as to chronologically anchor it (such as memories of 9/11), many of our memories are subject to the temporal distortions of the telescope effect. Finally, there is the common distortion of boundary extension, in which certain objects are remembered from a wider-angle view than they were experienced, creating the impression that the places in which these objects were initially encounter were larger than they actually are (Intraub & Hoffman, 1992). What all of these effects have in common is the fact that they present the remembered content in a distorted way, in a way that does not preserve the original structure of the experienced event. However, I submit that the mental experiences described by these effects still constitute genuine cases of

remembering. If I go back to my childhood room, and I show you how my books and toys were arranged, I do so in virtue of the fact that I remember my room and my books and my toys, even if my memory does not preserve the exact structure of the original perception (that is, even if I remember my room as being larger than it is). Consequently, the spatial and temporal distortions exemplified by these ordinary effects show us that there are genuine cases of remembering that do not preserve the structure of the original perception but are not mere imaginations (see Debus, 2007; Sutton, 2010).

Some might defend Russell's view that false memories are not the product of memory by arguing that these effects are actually the result of two different processes: one in which memory accurately recalls preserved information, and another one in which imagination distorts the structure of the perceptual components with which this information is conveyed. Thus, even cases of observer memories—where recalled information presents to our mind from the perspective of an observer different from us—would still count as remembering, insofar as memory effectively succeeds in recovering the preserved information. The distortion is merely a further effect of imagination. There are two problems with this defense. On the one hand, distortion is a matter of degree. If we allow small variations, like changes in perspective or shifts in the angle of our viewpoint, why can't we say the same of relatively larger variations, like the color of a remembered object or the gender of a remembered person? And if we admit those, why can't we say that larger variations—say, remembering pictures of a plane crash as footage—also count as memories? If we want to say that only *some* distorted memories count as genuine cases of remembering, the burden is on the objector to come up with a non-arbitrary way to tell the difference. On the other hand, this defense is threatened by

the possibility that far from being an exception, memory distortions may actually be the rule. Consider, for example, the fact that the temporal structure of our memories is normally distorted: remembering an event (almost) never takes the same amount of time it took us to experience it. Thus, if we say that imagination is involved in disarranging the temporal components of our memories, then it looks as though imagination may be involved in almost every recollection (I make a stronger case for this claim below). However, in saying that imagination is permanently involved in the exercise of memory, we risk blurring the distinction between these distinct faculties. The trick is to understand how the two can be different faculties even if some of their mechanisms overlap. This—as I suggest below—requires understanding the mechanisms of remembering differently.

Most philosophers are not persuaded by Russell's approach. They find the alternative strategy of saying that memorial effects like field-observer or boundary extension are indeed produced by memory, not by imagination, more appealing. Nonetheless, since the occurrence of these distortions shows that memory failed to preserve the structure of the original experience, then these particular recollections should be considered the product of a *malfunctioning* memory. Given that the function of memory is to accurately preserve the structure of the original perceptual representation, when it delivers reproductions of such perceptions with the wrong structure, memory is simply failing to perform its function. In summary, from this alternative perspective, all false and distorted memories are indeed the effects of our memory system, they just so happen to be produced by a failure of its mechanism.

The main problem with saying that false memories are a failure of memory is that it forces us to accept that we have a memory system that regularly and systematically

malfunctions. Evidence gathered by cognitive scientists in the last four decades makes it clear that false and distorted memories are a common occurrence in our daily lives. One of the most widely used experimental paradigms for false memory research is known as the Deese-Roediger-McDermott (DRM) paradigm. This paradigm consists of showing an individual a list of either perceptually or semantically related words (e.g. *tired, bed, awake, rest, dream, night, blanket, doze, slumber, snore, pillow, peace, yawn, drowsy*) that are associated to a non-presented lure (e.g. *sleep*). Subsequently, participants perform an old-new recognition task—that is, a task in which they have to say whether the word they are seeing is "old" (i.e. it was on the study list) or "new" (i.e. it wasn't on the study list)—when shown a list of words that include some of the previously presented words (e.g. *bed*), some non-presented non-related words (e.g. *hamburger*), and some non-presented related words or 'lures' (e.g. *sleep*). In general, false recognition of semantically and perceptually associated lures is quite high. Roediger and McDermott (1995) reported that participants falsely remembered critical lures 55% of the time—the exact same recall rate for words presented in the middle of the list! Similar effects have been reported in recognition tests of previously studied lists (Underwood, 1965), with participants falsely recognizing both synonyms and antonyms of previously studied words as having been in the list up to 30% of the time. Importantly, when the recognition list involves semantically related words, the false alarm rate can reach up to 70% (Payne et al., 1996). Finally, semantic intrusions have also been reported in experimental paradigms using sentences, showing that, under certain conditions, participants may report having heard an entire sentence that was not included in the original study set (Bransford et al., 1972).

In addition to words and sentences, psychologists have shown that people are prone to misremember perceptual details of previously witnessed events, and even entire events that never happened in their lives, but which people tend to recall as though they did. One of the most celebrated studies in eye witness suggestibility was conducted by Loftus in 1975. This study pioneers the use of the *misinformation paradigm*, which shows that people tend to report false memories when they receive misleading information during recall. Loftus presented participants with color slides depicting a car accident. The slides showed a car failing to stop at a traffic sign. Half of the participants were shown a slide with a 'stop' sign while the other half were shown a slide with a 'yield' sign. Twenty minutes after the slide show, participants received a 20-question interview, with the 17th question being the critical one. Half of the subjects that were shown the slide with the stop sign were asked if the car had failed to stop at the 'stop' sign, whereas the other half of that group were asked if the car had failed to stop at the 'yield' sign (the same occurred with the subjects that were shown the yield sign). The results were striking: on average, participants were unable to discriminate the correct answer. Even when subjects were told, before receiving the interview, that some of the questions may have stated the traffic sign incorrectly, participants still chose the correct sign only 43% of the time—versus 67% for those who were not misled (a remarkable result in itself, as it implies that participants chose the wrong answer 33% of the time even with no misinformation at all). In a follow up study varying the lag time between the stimulus and the misleading interview, Loftus and collaborators (1978) discovered that when the interview was administered 20 minutes after witnessing the event,

participants were correct about 40% of the time, but if the interview is administered one week after witnessing the event, the rate dropped to 18%.

Psychologists have also tested the misinformation paradigm using real-life autobiographical material, showing that people can misremember entire events that did not happen in their lives as though they did. Loftus and Pickrell (1995) showed that up to 25% of study participants would falsely remember having been lost in a shopping mall when they were children if they receive misleading information during suggestive interviews. Hyman (1995) showed the same effect for more unusual—although not implausible—events, like having been hospitalized or having had a party with clowns. More recently, Lindsay et al. (2004) used a variation of the misinformation paradigm involving doctored photographs. After seeing the photographs, 56% of participants falsely recalled experiencing an event (e.g., taking a trip in an air balloon) that they actually never experienced. The effects of the misinformation paradigm are related to those of the so-called *imagination inflation* effect, which shows that people tend to falsely remember an event as a result of having imagined what it would have been like to experience it prior to being asked to recall it. In one of the most telling demonstrations of this effect, Garry and collaborators (1996) randomly divided their participants into two groups. All participants were told that they were taking part in a study measuring their capacity to imagine counterfactual events with as much detail as possible. Two weeks later, the same participants were called back, this time to participate in an autobiographical recognition test. They were asked to state how sure they were that certain events did not happen in their lives. Half of the subjects were presented with events they had previously imagined; the other half were only presented with novel

events. Surprisingly, the confidence ratings for events that were previously imagined were substantially lower than for those that were new. In other words, participants were more confident saying that events they did not previously imagine definitely did not happen in their lifetime than they were saying that the events they had previously imagined definitely did not happen. Importantly, participants did not remember having imagined any of the critical events before (a phenomenon known as *source amnesia*).

So far, I have presented evidence suggesting that false memories are a normal occurrence in healthy individuals. Another piece of evidence in favor of this claim comes from studies probing the effects on memory-related tasks in subjects with episodic memory deficits. In a pioneer study, Schacter and collaborators (1996) endeavored to find out whether individuals with selectively impaired memory were more prone to misremembering than healthy subjects, a result that would lend credence to the view that false and distorted memories are the product of a faulty memory. They used the DRM paradigm on subjects with amnesia caused by medial-temporal lobe accidents. They discovered that individuals with amnesia showed significantly reduced false recognition of semantically and perceptually associated lures. Even though they were less accurate than controls overall, amnesiacs were significantly less likely to produce false alarms than controls (Melo et al., 1999; Ciaramelli et al., 2006). Other studies using visual shapes have revealed equivalent effects in amnesiacs, showing that the number of pictorial memory intrusions is significantly reduced relative to healthy controls (Koutstaal et al., 1999). Finally, similar studies with patients in the early stages of Alzheimer's disease—a neuropathology that usually begins at the medial-temporal lobes—have shown that, compared with age-matched controls, individuals with

Alzheimer's also present reduced false recognition rates (Balota et al., 1999; Budson 2003).

Taken together, these—and many similar—studies suggest several conclusions. First, they tell us that false and distorted memories are a common phenomenon. Second, they suggest that even though many of our ordinary memory experiences may be false recollections, entertaining those false memories not only does not affect us: we do not even notice if they are false memories. Third, the fact that individuals with memory-related pathologies tend to entertain less false and distorted memories than normal subjects, suggests that some degree of memory distortion may be non-pathological, and perhaps even beneficial. These studies also show that not all kinds of information are susceptible to being misremembered or are likely to implant false memories. False and distorted memories have an air of plausibility to them. One may falsely remember having experienced an episode from a distance, as it occurs with observer memories. Even though this in and of itself is strange, it may not be altogether implausible if one had been able to adopt someone else's point of view at the time. Similarly, having been lost in a shopping mall as a child, having a neighbor call 911 to complain about the noise, or having seen the word 'sleep' in a list of sleep-related words, are all plausible things that could have happened, at least in the sense of being more plausible than being abducted by talking unicorns, having the Beatles play at your sixth birthday party, or having seen the word 'multiplication' or 'vomit' in a list that otherwise contains words semantically related to fruits (Psychologists call this feature *schema-consistency*, meaning that false memories are consistent with schematic forms of the events they falsely portray. I discuss this issue further below). So the question is: how can the traditional view defend the

claim that false memories are the product of a malfunctioning faculty in the face of their pervasiveness and regularity? Furthermore: why would we have a cognitive system that malfunctions so constantly and so systematically? The answer may be that the function of memory is *not* to reproduce previous perceptions with the fidelity demanded by the traditional view. But then, how can we safeguard the obvious intuition that we use our memory to remember while accounting for the prevalence and regularity of our false memories?

In what follows I offer an answer to these questions. But first, let me summarize the main points of this first section. According to the traditional view, the function of memory is to remember, and remembering is thought of as the act of reproducing previous experiences with mental representations that preserve the structure of the original perceptions. As such, false memories occur when you deploy a mental representation whose content, although it is experienced as a memory, does not preserve the structure of any previous perceptual experience. From the traditional point of view, a false memory is understood as either the product of a cognitive mechanism other than memory (imagination) or as the product of memory malfunction. I argued that the first alternative forces us to say that some bona fide cases of remembering are not memories at all. The second route is also undesirable, for it makes it hard to understand why evolution would have favored a cognitive mechanism that malfunctions so regularly. In order to avoid these unwanted consequences, I argue we need to reconsider the way in which we think about cognitive functions.

3. Thinking about cognitive functions

The conclusion of the preceding discussion lends itself to an obvious objection. Even though most biological systems—cognitive and otherwise—tend to perform their functions regularly, there are many traits that malfunction frequently. However, functions are not statistically determined; the function of a system isn't necessarily equivalent to the role the system usually performs. A system may perform its normal function only rarely, when very specific conditions obtain—presumably the very conditions under which it contributed to the reproductive success of the organism's ancestors (Millikan, 1984)—but such infrequency does not license us to say that the system's function is not its normal one. Consequently, the fact that memory frequently malfunctions does not imply that it is not for remembering.

I think this objection can be put to rest, however. Consider two classic examples of allegedly frequently malfunctioning biological systems. Meerkats evolved an alarm call system for identifying predators. Nowadays, with fewer predators, their alarm system tends to produce more false alarms on average than it did in the past. However, the relative frequency at which it misfires does not take away from the fact that, given the circumstances in which it evolved, it was a highly reliable indicator of predators[22]. As a result, one can still say that the function of the system is to identify predators even if it regularly malfunctions. The second example is closer to home. Humans evolved particular cognitive strategies known as *heuristics* for quickly assessing the probability of certain events happening. One such strategy is the availability heuristic, according to which the relative facility with which information is made available to our conscious experience influences our perception of how probable an event may be. This is because

[22] The relationship between false alarms and predator population needn't be linear. It may be possible that false alarms also increase when there is an excessive number of predators. Still, the point I am about to make holds even in this hypothetical situation.

conscious availability tends to track frequency, which in turn tends to track probability. However, nowadays we face many situations in which the most probable event is not necessarily the one we have experienced as more frequent. As a result, when facing these situations, most people's judgments tend to align with our evolved heuristics, and thus produce the wrong judgments. But the fact that our judgments of probability so frequently lead us astray does not speak against the claim that the function of the cognitive systems with which we produce such judgments is to track the probability of events happening.

There are two crucial differences between these cases and memory which speak against taking false memories merely as instances of a frequently malfunctioning system. First, whereas in the case of the alarm call and the heuristic systems there is an identifiable change in circumstances responsible for the shift in the system's reliability, the same does not seem to happen with our memory system. The frequency of false alarms in the meerkat's alarm system varies as a function of the proportion of predators in the environment relative to other non-threatening objects that could trigger it. Likewise, in environments in which the most probable events are also those experienced as being the most frequent, our heuristic systems are reliable. When the most probable events aren't those experienced as the most frequent, our heuristic systems are not reliable. This change in circumstances is not apparent when it comes to the case of memory. The circumstances under which our memory system evolved are *not* significantly different from the circumstances in which we currently deploy it—or at least there is no particular circumstantial change that would have made memory evidently less reliable than it supposedly was. In fact, there may not be a particular feature of our

ancestor's environment whose change relative to our current environment shifted the conditions of reliability of our memory system.

The second crucial difference between these systems and memory is that the functional explanations of both the alarm call and the heuristic systems assume, rightly, that their purportedly correct functioning—identifying predators or probable events—is more beneficial than their malfunctioning. But the same assumption isn't warranted for the case of memory, as it is not obvious that a faithful reproduction of an experienced event is more beneficial than a relatively distorted reconstruction. Indeed, the exact opposite claim may actually be closer to the truth—namely that low rather than high fidelity in memory may provide us with a selective advantage. Psychologists Suddendorf and Corballis (2007) argue that memory systems contribute to the organism's fitness insofar as they allow it to recast knowledge of particular events that happened in the past in order to foresee what may happen in the future. They review a wide array of comparative studies in both human and non-human animals, and they conclude that successful anticipatory behavior is correlated with the degree of flexibility one's memory has to rearrange stored information. However, they also show that this flexibility demands relatively imprecise memory representations. This last idea, first proposed by Bartlett in 1932, is rapidly gaining popularity in contemporary cognitive science. Cognitive neuroscientists Daniel Schacter and Donna Addis (2007) suggest that considering the conditions upon which cognitive organisms like us live, encoding relatively sketchy or "gist-like representations" of previous experiences may be more advantageous. The thought is that, since we live in an informationally rich and constantly changing environment, and since there is a strict limit to the informational load we can

operate with at each time, both literal encoding and recall may become burdensome and risky (Bartlett 1932, 204). So, in order to respond quickly while saving storage space, our brains opt for a rather schematic way of encoding information.[23]

This line of thought lends itself to a tempting but ultimately wrong interpretation. According to this interpretation, the fidelity of memory would be cost effective, in the sense that in informationally rich environments faithful encoding becomes too costly, so memory opts for more gist-like representations of experienced events. In contrast, in informationally poor environments, memory can afford higher fidelity, so it encodes memories in ways that produce less distortion during recall. This possibility would preserve the view that memory is indeed for remembering, as memory distortions are simply the side effect of the elevated cost of high fidelity.[24] The problem with this interpretation is that it is hard to make sense of many of the experimental results discussed above in which false and distorted memories occurred in informationally poor environments where the stakes are pretty low—paradigmatically in psychology labs in which only short movies or brief word lists are presented. The amount of information

[23] A neat piece of evidence in support of the claim that distorted rather than faithful memory representations are more advantageous, would be to find out whether people who experience no memory distortion exhibit behaviors that are clearly less advantageous than those exhibited by people who normally experience memory distortions. There are several reasons why this piece of evidence is hard to gather in practice. For one, as I mentioned, false and distorted memories are prevalent and pervasive, to the extent that everybody seems susceptible to experience them. A longitudinal study comparing two groups in these two conditions would probably be impossible. An attractive alternative would be looking at specific populations. There are at least two possible populations from which to draw samples for a longitudinal study testing this hypothesis: patients with retrograde amnesia and individuals with hyperthymestic syndrome, a condition in which individuals appear unable to forget episodic details of their day-to-day lives. Multiple studies have been conducted with amnesic patients which, unsurprisingly, have a hard time getting around in the world. Some of these results will be covered shortly. But the second population remains understudied—partly because hyperthymestic syndrome is not only recent but also controversial. This, I believe, is a fecund line for future research.

[24] Contrast this interpretation with the case of the meerkat's alarm system. In the case of the meerkats, it looks as though the cost of a false alarm is significantly lower than the cost of missing a predator. Similarly, under this interpretation, the cost of producing distortions is significantly lower that the cost of encoding experiences with high fidelity.

participants experience in these environments is significantly lower than the amount of information one normally experiences in every-day situations. Therefore, since memory distortions are so general and frequent in these informationally poor environments, the idea of memory's fidelity as a function of the environment's informational richness loses its footing.

A better alternative, I contend, is to interpret the frequency of memory distortions not as a cost the system has to pay in exchange for efficacy, but rather as the beneficial byproduct of a mechanism that is actually doing something else. But, what could memory be for if not remembering? The answer to this question, I believe, requires a change of perspective as to how to determine the function of a cognitive system such as memory. As I mentioned, the traditional view follows a content-based approach, which consists of looking at the way in which the information processed by a cognitive function is experienced by the subject, and then going on to say that the mechanism responsible for those experiences must be there for that purpose. Thus, since memories are experienced as reproductions of previous experiences, the content-based approach recommends thinking of memory's function as that of reproducing previous experiences. However, there is no principled reason for us to pursue this approach. The way in which a particular mental content is experienced by us is orthogonal to the purpose of the system that is responsible for providing us with such an experience. Mental contents and purposes may or may not coincide, but there is no necessary connection between them. Perception, for instance, may be for guiding action (Noe, 2004), even if the content one is conscious of when perceiving is not experienced as such. Similarly, in the case of memory, all we know is that the way the contents of our memories are experienced by us is something

that memory *does*—and, as such, it is something one should expect to be accounted for once we have a adequate functional analysis of its operations—but it need not be what memory is *for*.

In order to find out what memory may actually be for, I suggest we look at it in terms of its contribution to cognitive organisms such as ourselves. Philosophers typically distinguish two general approaches to understand how a system contributes to an organism. The first approach is *etiological*. It consists in analyzing the function of a system in terms of its evolutionary history (e.g., Millikan, 1984). The second approach, which we may call the *role function* approach, consists in analyzing what something is for in terms of its contribution to the organism qua biological system (Cummins, 1975; 1983). Although both approaches are complementary—what one says about a system's contribution to the containing organism must also make evolutionary sense (Godfrey-Smith, 1994)—my objective here aligns more with the second approach. Specifically, I want to consider memory as a cognitive mechanism whose function is determined by the activities with which it contributes to the overall goals of the containing cognitive organism (see Machamer et al., 2000; Craver, 2000). As such, if we think of cognitive organisms like humans in terms of our overarching goals—survival, for instance—memory's "mechanistic role function" (Craver, 2000: 61) would be its contribution toward that goal. Thus understood, the task we face now is that of determining what the mechanistic role function of our memory system may be.

Explanations of mechanistic role functions are thought to be hierarchical. This is not at all a new idea. It is the backbone of what functionalist philosophers called "functional analysis"; that is, the idea that the mind could be decomposed into

hierarchical computational levels until it bottoms out at the ultimate level of implementation. What is novel about the functional mechanistic role approach, however, is that it tries to spell out what it means for a *mechanism* to carry out a function at a particular level of the hierarchy. Thus, in order to determine the mechanistic role function of a particular cognitive system, *S*, one needs, first, to determine the way in which the mechanisms of the immediately lower level contribute to *S*'s functioning, and, second, how *S* contributes to the proper functioning of the mechanism at the immediately superior level. However, in so doing, we may end up with something that the traditional functional analysis did not foresee (except, perhaps, for Lycan, 1988: 31-32): we may have been just wrong about the mind. More precisely: we may end up discovering that the mechanistic role function of a certain cognitive system does not square with our folk psychological characterization of its alleged function. Successful functional analyses may reveal that the hierarchical organization of the mechanisms carrying out the computations underlying our mental processes might not constitute a smooth linear transition from mind talk to brain talk. Bechtel makes this point quite clearly when he says that,

> The decomposition of mental life into activities such as perception, memory, and decision making provided the takeoff point for the development of cognitive accounts of the mind. […] One of the most potent contributions of the attempt to localize memory and other activities in the brain may be the realization that such decompositions may not reflect how the brain organizes its functioning. When a variety of mental operations have been convincingly identified and localized in the appropriate brain areas, it may turn out that the characterization of the operations are orthogonal to these long-standing categories of mental phenomena. (Bechtel, 2008: 82).

I believe this is precisely what happens with memory. According to its mechanistic role, remembering does not seem to be what memory is for, as remembering appears to be a sub-routine of a larger and more complex operation—or so I argue in the next section.

To review, in the first section I argued that the traditional view cannot account for recent evidence in the cognitive psychology and neuroscience of false memory, unless we are willing to accept one of two unappealing consequences: either that some genuine cases of remembering are not the product of memory or that it is adaptive to have a memory system that frequently and regularly malfunctions. This last alternative left open the possibility of thinking of memory as a system that indeed evolved to perform the function of remembering, but that like other adaptive traits that frequently malfunction, under new circumstances it regularly produces false and distorted memories. In this second section I argued against this possibility by noting, first, that the relevant circumstances under which our memory system evolved are not significantly different from those in which we currently deploy it and, second, that empirical evidence suggests that some degree of memory distortion appears to be more adaptive than faithful reproductions of previous events. I suggested, then, that our failed attempts to safeguard the traditional view may be rooted in a false methodological assumption: the content-based approach. I recommended pursuing an alternative approach—the mechanistic-role function approach—according to which the function of a cognitive system is determined by the way in which the mechanism that instantiates the system contributes to the overall goals of the containing organism. In the next and final section I explain how, when seen from the point of view of its mechanistic-role function, memory is not for remembering, but rather for recombining pieces of memory traces into mental representations of what

could have happened in our past in the service of creating optimal predictions of what could happen in our future. This perspective will dispel the unwanted consequences unveiled in the first section by explaining why ordinary cases of false and distorted memories aren't the product of a faulty mechanism, but rather the appropriate result of the same underlying process that produces veridical memories.

4. Remembering what could have happened.

As I mentioned above, there are two parts to understanding the mechanistic role function of a system: one needs to understand how the mechanisms that *compose* such a system work, and also how the system contributes to the functioning of the larger system that *contains* it. Cognitive neuroscience has managed to give a relatively accurate depiction of the mechanisms underlying recollection. I want to show now, by way of an example, the role of such mechanisms during an instance of ordinary false recollection. The purpose of this illustration is to make a case for the thesis that the mechanisms underlying the recollection of *some* false memories are identical to those underlying veridical remembering. From this illustration I will draw three morals regarding the nature of such mechanisms. I suggest that these features only make sense if we consider memory as a subroutine of a larger system whose function is not to reproduce past experiences, but rather to recombine them in order to entertain what I call *episodic counterfactual thoughts.*

The example I have in mind echoes one of the false memory paradigms mentioned above (Loftus, 1975). Suppose you are driving and you witness a car accident. A red Datsun fails to stop at a 'yield' sign and hits the truck next to you. The event takes

no more than eight seconds. During that brief period of time, your attention is shifting everywhere. A large percentage of that time your attention is focused on stimuli relevant for your survival: the front of the car, the incoming vehicle, the wheel, the brake pedal. A smaller percentage is probably allocated to some perceptual details that call your attention on the basis of their biological relevance: the face of the driver, the color of the car, the sound of the tires on the road. The remaining percentage—probably a very small percentage—may be allocated to other details, like the words written on the traffic sign, or its particular shape. Anything else that was not attended was not processed by working memory. So, in addition to the unattended details, those that were not rehearsed in working memory because they exceeded its informational quota would not have been consolidated in long term memory. Only a small portion of the information that made it all the way through would be effectively encoded in long term memory, with its sensory components dispersed over the sensory cortex (viz., visual information about the event will be schematically stored in occipital areas, auditory information in dorsal temporal areas, semantic information in ventral temporal areas, etc.) This is the sense, I understand, in which Schacter and Addis (2007) talk of storing a gist-like representation.[25]

Now suppose that, thirty minutes after you witnessed the accident, a policeman calls and asks you to remember whether the red Datsun failed to stop at the stop sign. During recall, your pre-frontal cortex—aided by the medial temporal lobes and the

[25] "Storing" is a rather misleading term. What seems to occur when we encode information is the strengthening of neural connections due to the co-activation of different regions of the brain, particularly in the sensory cortices, the medial temporal lobe, the superior parietal cortex, and the lateral prefrontal cortex. During encoding, each of these regions performs a different function depending on the moment in which the information gets processed. A memory trace is the dispositional property these regions have to re-activate, when triggered by the right cue, in roughly the same pattern of activation they underwent during encoding. (See chapter 4).

superior temporal cortices (Shimamura, forthcoming; and chapter 4)—calls back the disaggregated sensory information stored in the neural connections engaged during the event perception (e.g., the color red, the sound of squealing tires, your bodily reaction, etc.). Since they are disaggregated, the act of recollection becomes the act of reconstructing disperse encoded information. However, given that not all the relevant information was effectively encoded, this reconstruction is more like trying to rebuild a dinosaur out of its fossilized remains (Neisser, 1967) than trying to put together a jigsaw puzzle of which you have all the pieces. Memory, therefore, must be the sort of mechanism that can not only bind together information that is distributed across neural networks, but must also be able to fill the gaps left by the missing pieces. Importantly, though, the process by means of which memory fills those gaps is not haphazard. On the contrary, recent work on computational models of cognition using parallel distributed networks has shown that, by inserting probabilistic constraints based on a prior optimization hypothesis constructed out of previous experiences with similar stimuli, one can replicate and even predict memory intrusions (for a classic attempt, McClelland, 1995; for a more sophisticated one, Steyvers et al., 2006). The idea—to go back to the example—is that since you have had many similar experiences in comparable situations—i.e., with comparable cars and comparable traffic signs—your perceptual system has grown accustomed to receiving very specific kinds of visual information during perceptual events just like this one. As a result, the connection weights among the nodes of the neural network have been attuned to the probability of seeing 'stop' signs more often than 'yield' signs—or, for that matter, than gigantic lollipops. Your perception, therefore, does not need to attend to every detail of the stop sign when

encoding, since due to previous similar experiences it can optimize the process by filling the missing pieces according to probabilistic rules—think of the schema consistency effect mentioned above. In turn, memory takes advantage of this very same mechanism, and instead of storing an altogether new copy of the optimized visual stimulus, it simply creates an index—presumably in the parahippocampal cortex (Nadel & Moscovitch, 1997)—that tells the brain which neural networks engaged during the original perceptual event need to be reactivated during recall. So it is no surprise that now that you are trying to remember the accident after having been cued by a related piece of auditory information that has been added to the process of recollection—namely the words "stop sign" said by the policeman—the reactivation of the sensory information brings to your mind an optimized mental representation of the event, which, thanks to your memory's proper functioning, fills the gap left by the unattended spot with what it finds to be more likely: a visual image of a 'stop' sign. You have definitively misremembered the event, there is no question about that, but your memory mechanism was working just fine.

There are three morals I would like to draw from this illustration. First, when I say that veridical and false memories are the workings of the same mechanism, I mean to imply that just as there are cases of false memories that are like cases of veridical memories, there are also cases of veridical memories that are like cases of false memories. Given that recollection is probabilistic in the sense suggested above, successful encoding just means increased probability of recall[26]. Therefore, a successful recall produced by reconstructing the optimal representation of an event given a cue

[26] This assertion is contentious but important. Memory traces or "engrams" do not have the ontological status of objects or events. They are dispositional properties of neural networks to elicit certain responses (see footnote 6). A similar idea can be found in the works of Semon (1909), Martin and Deutscher (1966), and more recently Tulving (2002).

would count as veridical recall even if some sensory details of the original stimulus were not attended. Attending to those details would increase the probability of successfully recalling them later on, of course, but since we cannot afford having every aspect of the world under the spotlight of attention, the next best solution is to have a system that can fill in the gaps with the optimal alternative it can come up with. Most of the time what you recall accurately depicts the witnessed event. Sometimes it doesn't. In both cases, however, the system is doing what it is supposed to do.

The second moral is that our memory system redeploys mechanisms that can be used for purposes other than recollection. As I pointed out, evidence shows that the same regions of the sensory cortex recruited during the perceptual processing of a particular event are later on reactivated during the recollection of the same event (e.g., Wheeler et al., 2000). This suggests that, depending on the cognitive task they are engaged in, these very regions are playing a different role. Similarly, the superior frontal gyrus, the superior parietal cortex and the hippocampal complex, all involved in episodic recollection, have shown to be actively engaged in several distinct cognitive tasks (Andersen et al., 2007). Indeed, the more we know about brain processes, the clearer it is that far from being an exception, massive redeployment of neural systems may actually be the rule (Anderson, 2007; 2010). Consequently, a successful account of our memory system needs to be consistent with the way in which its components are redeployed for other cognitive tasks.

Finally, the third and—for the purposes at hand—most relevant moral, is that our memory system employs a probabilistic strategy to recover information. This fact dovetails with one of the most interesting features unveiled by the research on false memory mentioned in the first part of this paper: most memory distortions are *schema*

consistent. When people misremember events—or mere details thereof—their false or distorted memories are usually about things that even though they did not happen—or did not happen exactly as they were misremembered—they, nonetheless, *could* have happened. This suggests that our memory system must be sensitive, not only to what indeed happened in our past, but also to what could have plausibly happened. How could this be? A reason for that, I suggest, is that the mechanisms underlying our episodic memory are contained within a larger system that supports episodic counterfactual thinking, that is, thoughts of what could happen in one's future as well as what could have happened in one's own past.

Some initial evidence in support of this claim comes from cognitive neuroscience. Current studies of autobiographical episodic recollection reveal a striking structural parallelism between the brain structures recruited for episodic memory and future projection. In particular, recent evidence suggests a relation between remembering one's past and imagining one's future. Neuropsychological studies have shown that deficits in autobiographical episodic remembering are associated with deficits in people's capacities to project themselves into the future. Evidence supporting this claim comes from research on amnesic subjects (Tulving, 1985; Klein et al., 2002; Hassabis et al., 2007), older adults with memory-related pathologies (Addis et al., 2009), patients with severe depression (Dickson & Bates, 2005; Williams et al., 1996), and subjects with schizophrenia (D'Argembeau et al., 2008). In addition, research in healthy individuals has shown that when certain phenomenonological features of our prospective thoughts are manipulated—e.g., vividness, emotional significance, etc.—the effects are very similar to those elicited by equivalent manipulations in autobiographical recollection

(D'Argembeau & Van der Linden, 2004; Szpunar & McDermott, 2008). Finally, research using neuroimaging techniques has revealed a *core brain network* that is engaged during autobiographical remembering and future projection (Schacter et al., 2007; Addis & Schacter, 2008; Addis et al., 2007). This core brain network involves the hippocampus, the posterior cingulated/retrosplenial cortex, the inferior parietal lobe, the medial prefrontal cortex, and the lateral temporal cortex. Importantly, this core brain network is *not* engaged when people are asked to imagine events that do not involve considerations of what could happen to them (as opposed to considerations of what could happen to other people) (Hassabis et al., 2007; Okuda et al., 2003; Szpunar et al., 2007). Not all kinds of imagining, therefore, seem to be reliant on the same brain network.

Further evidence in support of the claim that the same neural mechanisms engaged during episodic recollection are also recruited for thinking of what may happen, comes from recent behavioral and neurophysiological studies in non-human animals. Human-like episodic memory has been notoriously difficult to find in non-human animals. Surprisingly, however, recent studies in cognitive ethology and comparative psychology have come to a similar conclusion regarding the capacity of non-human animals to foresee future events. Although many animals exhibit future oriented behaviors, these are inflexible and most likely instinctual. Rats, for instance, continue to cache food in locations of the maze that have shown to damage the food in the past, as though they had little sense of what that location could do to new food in the future (McKenzie et al., 2005). Similarly, New Caledonian crows develop replenishment routines that are highly dependent on the regularities in food supply at different locations. However, they are incapable of breaking those routines in order to circumvent

environmental contingencies (Burke & Fulham, 2003). Finally, volumetric comparisons between ape and human brains confirm that critical regions of the brain core network—in particular, inferior pre-frontal cortex (BA 11 and 47) and the frontal pole (BA 10)[27]—are orders of magnitude larger in humans than in our closest non-human relatives (accounting for their encephalization quotient, a widely used ratio of brain weight to body weight. See Flinn et al., 2005).

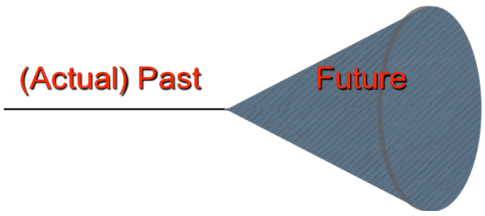
Taken together, these results lend strong credence to the view, suggested previously by Tulving (1985), that episodic autobiographical memory provides the basis for "mental time travel": our ability to mentally travel back in time to relive past experiences and to mentally project ourselves onto the future in order to anticipate what may come. In turn, these results suggest that the same mechanisms recruited for thinking about what happened in our past are also responsible for our thinking about what may happen in our future. And since thoughts of what may happen in the future are a species of thoughts of what may happen, then this is at least initial evidence to the effect that the same mechanisms responsible for our capacity to think about what happened in our past may also subserve our capacity to think of what may happen in general.

This particular view of episodic memory as "mental time travel" unveils an interesting asymmetry: whereas our recollection of an autobiographical memory is constrained by how the experienced facts actually were, prospective thought is not so constrained [Figure 4]. When we remember what happened to us on a particular occasion,

[27] This is an important piece of information for the inferior frontal and frontopolar cortex are preferentially engaged during episodic counterfactual thinking, i.e. our capacity to think about alternative outcomes to events that happened in our past (De Brigard et al., forthcoming).

what comes to mind is determined by what we lived; any deviation from what we lived allegedly constitutes a false memory. However, when we think of ourselves in the future, there is no lived experience against which to contrast the content of our mental representation, and as a result the scope of possibilities that can be brought to mind isn't determined by the facts. To put it simply: if this view of episodic memory as "mental time travel" is correct, there is only one way to mentally travel back to the past, but many ways to mentally travel into the future. Could it be possible that the same core network underlying recollection and prospection could also subserve episodic counterfactual thinking, i.e. thoughts about what could have happened in one's own life?

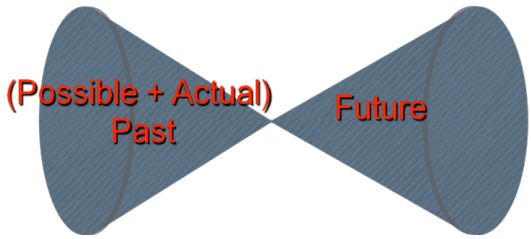
Figure 4:



To test this possibility, my collaborators and I conducted a study in which participants were presented with alternative outcomes to events they experienced in their lives while undergoing functional magnetic resonance imaging (De Brigard et al., forthcoming). In particular, using a variation on a paradigm previously employed for looking at the neural correlates of future thought, we asked participants to imagine what would have happened if the outcome of an experienced event would have been different than it was, either in a positive or a negative way. The results were striking. Thinking of

what could have happened in the past but did not recruited most of the same regions that are engaged during autobiographical episodic recollection. Specifically, it engaged regions in superior and medial frontal gyrus, left inferior parietal lobule, middle occipital gyrus, cyngulate gyrus, claustrum and putamen. Importantly, many of these regions belong to the aforementioned core brain network, suggesting that the network not only enables episodic recollection and future projection but also episodic counterfactual thinking. These results are of a piece with previous interpretations of the brain core network according to which its function is to enable us to perform cognitive tasks involving self-projection (Buckner & Carroll, 2007). But since thoughts of self-projection are also a species of episodic counterfactual thought, it may be more natural to think of the brain core network as being engaged in thoughts of what could happen to us either in the possible past—with our actual past being an actualized version of a possible past—or in the possible future [Figure 5].

Figure 5:



Where do we stand on the issue of the function of memory? The evidence reviewed suggests that the mechanisms underlying our episodic memory system are integrated within a larger system that supports thoughts of what could happen to us in the

future as well as what could have happened to us in the past. Remembering, therefore, should be understood merely as a particular operation of a system whose function is to enable us to entertain episodic counterfactual thoughts. This implies that it is a mistake to think that 'memory' refers to a system uniquely dedicated to reproducing the contents of previous experiences. What we normally call 'memory', and what many have tended to reify as an independent system, is a misnomer for a particular subset of operations carried out by a larger system that supports episodic counterfactual thoughts.

Thinking of memory in this mechanistic way can help us explain the phenomena that the traditional view could not accommodate. Veridical memories are produced by an underlying probabilistic mechanism. When we try to recall an event, this mechanism reconstructs an optimized mental representation from the encoded perceptual information according to probabilistic constraints dictated by previous experiences. As such, successfully encoded perceptual information is more likely to be remembered than information that wasn't successfully encoded. However, most of the time, due to informational limits on both perception and working memory, we fail to encode several informational details. Fortunately, memory's probabilistic nature fills in the missing information according to the very same optimization algorithms that it would have followed had the information actually been encoded. As a result, when unattended information deviates from what would have been its optimal reconstruction, the recollection of the event would likely yield a misrepresentation of what indeed happened. This is precisely what occurs with the boundary extension effect. We tend to experience middle-size objects from angles that allow us to fully see them. When we are presented with objects missing their boundaries and we fail to focus our attention in their missing

frontiers, we will tend to recall them from the point of view from which we would have normally experienced them, i.e. a wider angle. A similar explanation is available for the misinformation paradigm. During episodic retrieval, reconstructed memory traces are vulnerable to the conditions of recall. If you hear or see a cue while retrieving, that cue could influence the ongoing process of probabilistic reconstruction, so that it may come to fill the gap left by a tenuously encoded initial perception.

On the other hand, the fact that memory is an integral part of a larger system supporting episodic counterfactual thought also helps explain the fact that sometimes we remember things from an observer perspective. After all, observer memories are always perceived from the perspective one could have had—had one been an observer of the event. A contrived counterfactual, no doubt, but not implausible—especially not if we consider the fact that the brain core network is also engaged during tasks requiring theory of mind, i.e. adopting someone else's point of view (Buckner & Carroll, 2007). This model also helps explain why amnestic individuals tend to produce proportionately fewer distortions and false memories than their healthier counterparts. If my picture of the neuroanatomical underpinnings of the memory mechanism is roughly correct, the anterior regions of the hippocampus would permit the flexible recombination of the perceptual components encoded in the sensory cortex. As a result, an atrophy in the medial-temporal lobes would render binding those pieces together impossible, which in turn would make it impossible to reconstruct both veridical and distorted memories. From this perspective, then, amnestics haven't lost their memory, but their capacity to bind together previously acquired information into plausible episodes that may or may not have happened in the past or that may or may not happen in the future. Finally, the model I am suggesting

helps explain the phenomenon of imagination inflation. The sorts of things that people tend to misremember after having imagined them are highly plausible. People misremember having seen a stop sign, but they don't misremember having dated a celebrity. And when they do—and only when they do—we may be able to say that their memory is malfunctioning (as it occurs in the case of confabulation), but the same conclusion isn't always warranted when it comes to more quotidian cases of false and distorted memories.

Let me conclude by summarizing the main points of this last section. After arguing against the necessity of pursuing a content-based approach to analyze the function of memory, I weighed in favor of a mechanistic-role function approach. According to this perspective, in order to understand the function of a cognitive system like memory, one needs to understand both the mechanisms of which it is composed as well as the larger system that contains it. I then showed how the mechanisms underlying memory behave according to probabilistic strategies for filling in incomplete memory traces following optimization constraints. As a result, many ordinary false and distorted memories are the appropriate product of the same underlying probabilistic mechanism that produces veridical memories. That the brain follows this strategy could be interpreted merely as a tradeoff it has to make in order to balance accuracy and efficiency. According to this interpretation, just as meerkats evolved an alarm system that traded on reliability for the sake of efficiency, memory could be seen as a system that trades on accuracy for the sake of efficiency by storing gist-like representations later to be filled in probabilistically. Consequently, memory could still be seen as being for remembering, as false and distorted memories are merely by-products of this tradeoff.

However, I argued against this interpretation by noting, first, that memory distortions seem to be inversely correlated with forecasting, which is presumably an adaptive trait. I then argued that people show equal rates of memory distortion in informationally rich environments as they show in informationally poor environments. Finally, I argued that memory is integrated into a larger cognitive system that allows us to entertain thoughts of what could happen to us in the future by way of flexibly recombining perceptual components of previous experiences into episodic counterfactual thoughts. As a result, memory distortions are not the undesired by-products of an imperfect system, but rather the natural effects of a healthy cognitive system that is constantly using stored information to rehearse alternative pasts in order to discipline our thoughts about what may happen in the future. Thus, from a mechanistic-role point of view, remembering becomes simply a particular operation of a system whose actual function is to enable us to entertain episodic counterfactual thoughts. Lastly, I contend that this view, if correct, should dispel the unwanted consequences produced by our commitment to the traditional view of the function of memory. In particular, it permits us to keep the intuition that when we falsely remember we are actually exercising our memory while allowing us to make sense of the pervasiveness of false and distorted memories without having to consider them as the product of a malfunctioning mechanism.

References

Addis, D.R., Wong, A.T., & Schacter, D.L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45: 1363-1377.

Addis, D.R., & Schacter, D.L. (2008). Effects of detail and temporal distance of past and future events on the engagements of a common neural network. *Hippocampus*, 18: 227-237.

Addis, D. R.**,** Sacchetti, D. C., Ally, B. A., Budson, A. E., Schacter, D. L.(2009). Episodic simulation of future events is impaired in mild Alzheimer's disease. *Neuropsychologia, 47,* 2660-2671.

Adlam, A.; Vargha-Khadem, F., Mishkin, M, & de Haan, M. (2005). Deferred Imitation of Action Sequences in Developmental Amnesia. *Journal of Cognitive Neuroscience*. 17(2): 240-248.

Andersen, P.; Bruland, H. & Kaada, B. (1961). Activation of the Field CAI of the Hippocampus by Septal Stimulation. *Acta Physiologica Scandinavica*. 51 (1): 29-40.

Anderson, M. (2007). The massive redeployment hypothesis and the functional topography of the brain. *Philosophical Psychology*, 21(2): 143-174.

Anderson, M. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, 33(4): 245–313.

Audi, R. (1998), *Epistemology*, London: Routledge.

Bach, K. (1997). Do Belief Reports Report Beliefs? *Pacific Philosophical Quarterly*. 78: 215–41.

Balaguer, M. (1998). Attitudes Without Propositions. *Philosophy and Phenomenological Research*. 58 (4), 805-826.

Balota, D.A., Cortese, M., Duchek, J.M., Adams, D., Roediger, H.L., McDermott, K., & Yerys, B. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer type. *Cognitive Neuropsychology*, 15, 361-384.

Barr, R., & Hayne, H. (1996). The Effect of Event Structure on Imitation in Infancy: Practice Makes Perfect? *Infant Behavior and Development.* 19: 253-257.

Barr, R., Rovee-Collier, C. K., & Campanella, J. (2005). Retrieval protracts deferred imitation by 6-month-olds. *Infancy*, 7, 263 – 284.

Barsalou, L. W. (1999). Perception of Perceptual Symbols. Behavioral and Brain Sciences. 22(4): 637-660.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

Berntsen, D. (2009). *Involuntary autobiographical memories. An introduction to the unbidden past.* Cambridge: Cambridge University Press.

Bechtel, W. (2008) *Mental Mechanisms.* NY: Routledge, Psychology Press

Benjamin, B.S. (1956). Remembering. *Mind.* 65: 312-331.

Bergson, H. (1908). *Matter and Memory*. New York: Zone Books

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

Bernecker, S. (2010). *Memory*. Oxford: Oxford University Press.

Berryhill, M. E., Phuong, L., Picasso, L., Cabeza, R., & Olson, I. R. (2007). Parietal lobe and episodic memory: Bilateral damage causes impaired free recall of autobiographical memory. *Journal of Neuroscience, 27,* 14,415–14,423.

Bisiach, E. & Luzzatti, C. (1978). Unilateral Neglect of Representational Space. *Cortex*, 14, 129-33.

Bliss T.V., & Lomo T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology.* 232(2): 331-56.

Borg, E. (2002). Pointing at Jack, Talking about Jill: Understanding Deferred Uses of Demonstrative Pronouns. *Mind and Language.* 17(5): 489-512.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory.* New York: Oxford University Press.

Bransford, J.D., Barclay, J.R., & Franks, J.J. (1972). Sentence memory: A constructive versus interpretative approach. *Cognitive Psychology.* 3, 193-209.

Broad, C.D. (1925), *The Mind and its Place in Nature*. London: Routledge and Kegan Paul.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*, 49–57.

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition, 51,* 251–261.

Burke, D. & Fulham, B. J. (2003). An evolved spatial memory bias in a nectar-feeding bird. *Animal Behavior* 66:695–701.

Cabeza, R. (2008). Role of parietal regions in episodic memory retrieval: The dual attentional processes hypothesis. *Neuropsychologia*, 46, 1813-1827

Campbell, J. (2002). *Consciousness and Reference*. Oxford: Oxford University Press.

Cartwrtight, N. (1983). *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Chalmers, D. (1996). Syntactic Transformations on Distributed Representations. *Connection Science*. 2: 53- 62.

Ciaramelli, E., Ghetti, S., Frattarelli, M., & L'adavas, E. (2006). When true memory availability promotes false memory: Evidence from confabulating patients. *Neurophychologia, 44*, 1866-1877

Clarke, S. (2001). Defensible Territory for Entity Realism. *British Journal for the Philosophy of Science.* 52: 701-722.

Corbetta, M. & Shulman, G. L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3, 201-215.

Crane, T. (2009). Is perception a Propositional Attitude? *Philosophical Quarterly.* 59(236): 452-469.

Craver, C. (2001). Role Functions, Mechanisms, and Hierarchy. *Philosophy of Science*. 68: 53-74.

Craver, C. (2002). Interlevel Experiments and Multilevel Mechanisms in the Neuroscience of Memory. *Philosophy of Science.* Supplemental 69: S83-S97.

Craver, C. (2003). The Making of a Memory Mechanism. *Journal of the History of Biology.* 36: 153-195.

C.F. Craver (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience.* Clarendon Press: Oxford.

Crimmins, M., & Perry, J. (1989). The Prince and The Phone Booth: Reporting Puzzling Beliefs. *Journal of Philosophy.* 86(12): 685-711.

Crombag, H. F. M., Wagenaar, W. A., and van Koppen, P. J. (1996). Crashing memories and the problem of 'source monitoring. *Applied Cognitive Psychology, 10*, 95-104.

Cummins, R. (1975). "Functional Analysis." *Journal of Philosophy* 72: 741.765.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.

Danker, J. F. & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin. 136*: 87-102

D'Argembeau, A., & van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Consciousness & Cognition*, *13*, 844–858.

D'Argembeau, A., Raffard, S., & van der Linden, M. (2008). Remembering the past and imagining the future in schizophrenia. *Journal of Abnormal Psychology*, *117*, 247–251.

De Brigard, F. (2007). What was I thinking? An essay on the nature of propositional attitudes. MA Thesis: UNC.

De Brigard, F. & Prinz, J. (2010). Attention and Consciousnes. *Wires Interdisciplinary Reviews.* 1(1): 51-59.

De Brigard, F., Addis, D.R., Ford, J.H., Schacter, D.L., & Giovanello, K.S. (Forthcoming). Remembering what could have happened. The cognitive neuroscience of episodic memory and episodic counterfactual thinking.

Descartes, R. (1991), *The Philosophical Writings of Descartes, vol. III: correspondence*, J. Cottingham, R. Stoothoff, D. Murdoch, and A. Kenny (trans.). Cambridge: Cambridge University Press.

Debus, D. (2007). Perspectives on the Past: A Study of the Spatial Perspectival Characteristics of Recollective Memories. *Mind and Language* 22 (2):173-206.

Dickson, J. M., & Bates, G. W. (2005). Influence of repression on autobiographical memories and expectations of the future. *Australian Journal of Psychology*, *57*, 20–27.

Dretske, F. (1969). Seeing and Knowing. Chicago: Chicago University Press.

Dretske, F. (1982). Knowledge and the flow of information. Cambridge: MIT Press

Duncan, C.P. (1949) The retroactive effect of electroshock on learning. *J. Comp. Physiol. Psychol.* 42:32–44

Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.

Evans, N. (2007). Standing up your mind: Remembering in Dalabon. In: Amberber, M. (ed). *The Language of Memory in a Cross-linguistic perspective.* John Benjamins: Amsterdam.

Evans, N. & Levinson, S.C. (2009). The Myth of Language Universals: Language Diversity and its Importance for Cognitive Science. *Behavioral and Brain Sciences.* 32: 429-492.

Fellini L., Florian C., Courtey J., Roullet P. (2009) Pharmacological intervention of hippocampal CA3 NMDA receptors impairs acquisition and long-term memory retrieval of spatial pattern completion task. *Learning & Memory* 16, 387-394.

Fernandes, M.A. & Moscovitch, M. (2000). Divided attention and memory: Evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General*, 129, 155-176.

Fernandes, M. A., Moscovitch, M., Ziegler, M., & Grady, C. (2005). Brain regions associated with successful and unsuccessful retrieval of verbal episodic memory under divided attention. *Neuropsychologia*, 43, 1115-1127.

Fink, M. (1990). How does convulsive therapy work? *Neuropsychopharmacology*. 3(2): 83-7

Fodor, J. 1974. "Special Sciences". Reprinted in: Fodor, 1981.

Fodor, J. (1978). Propositional Attitudes. Reprinted in: Fodor, 1981

Fodor, J. (1981). *Representations*. Cambridge, MA: MIT Press

Fodor, J. 1985. "Fodor's Guide to Mental Representation". *Mind*. Spring: 66-97.

Flexner J.B, Flexner L.B, Stellar E. (1963). Memory in mice as affected by intracerebral puromycin. *Science.* 141:57–59.

Fraser, L.M., O'Carroll, R.E., Ebmeier, K.P. (2008). The effect of electroconvulsive therapy on autobiographical memory: a systematic review. J. ECT 24, 10–17.

Flinn, M. V., Geary, D. C. & Ward, C. V. (2005) Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence. *Evolution and Human Behavior* 26:10–46.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10, 680-698.

Frankland, P.W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nat. Rev. Neurosci.* 6:119 -130

Furlong, E.J. (1948), 'Memory', *Mind*, 57: 16–44.

Garry, M., Manning, C.G., Loftus, E.F., & Sherman, S.J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*, 208-214.

Gibson, J.J. (1979), *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin.

Gomulicki, B.R. (1953) *The Development and Present Status of the Trace Theory of Memory*, Cambridge University Press: New York.

Godfrey-Smith, P. (1994). A modern history theory of functions. *Nous* 28: 344-362.

Gottfried, J.A, Smith A.P., Rugg, M.D., Dolan, R.J. (2004) Remembrance of odors past: human olfactory cortex in cross-modal recognition memory. *Neuron* 42:687– 695.

Green, J.D. & W.R. Adey. (1956). Electrophysiological studies of hippocampal connections and excitability, Electroenceph. clin. Neurophysiol. 8.

Greenberg, D.L, Eacott M.J., Brechin D, & Rubin D.C. (2005). Visual memory loss and autobiographical amnesia: A case study. *Neuropsychologia.* 43(10):1493–1502

Hacking, I. (1983). *Representing and Intervening*. Cambridge: Cambridge University Press.

Halsey, R. & Chapanis, A. (1951). On the number of absolutely identifiable spectral hues. *Journal of the Optical Society of America*, 41, 1057-1058

Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature* 406(6792):147-50.

Hardt, O., Einarsson, E. Ö., & Nader, K. (2010). A Bridge over troubled water: Reconsolidation as a link between cognitive and neurotraditions. *Annual Review of Psychology*, 61, 141-167

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1726–1731.

Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: Age-related changes in deferred imitation. *Behavioral Neuroscience*, 114, 77 – 83.

Hazlett, A. (2010). The Myth of Factive Verbs. *Philosophy and Phenomenological Research* 80 (3):497-522.

Heil, J. (1978),'Traces of Things Past', *Philosophy of Science*, 45: 60–72.

Hitchcock, C. (1992). Causal Explanation and Scientific Realism. *Erkenntnis*. 37: 111-178.

Hobbes, T., 1651, *Leviathan*, in E. Curley (ed.), *Leviathan, with selected variants from the Latin edition of 1668*, Indianapolis: Hackett, 1994.

Hoefer, C. (2008). Introducing Nancy Cartwright's Philosophy of Science. In: Hartman, S.; Hoefer, C., & Bovens, L. (eds.) *Nancy Cartwright's Philosophy of Science.* Routledge: London.

Hoerl, C. & McCormack, T. (2004). Joint Reminiscing as Joint Attention to the Past. In N. Eilan, C. Hoerl, T. McCormack, and J. Roessler (eds), *Joint Attention: communication and other minds.* Oxford: Oxford University Press.

Hofweber, T. (2006). Inexpressible Properties and Propositions. In: *Oxford Studies in Metaphysics*. V. 2. D. Zimmerman (ed.) Oxford: Oxford University Press.

Hume, D. (1978). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hyman, I. E., Jr., Husband, T. H., & Billings, F. J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology, 9*, 181-197

Intraub, H., & Hoffman, J.E. (1992). Remembering scenes that were never seen: Reading and visual memory. *American Journal of Psychology, 105*, 101-114.

James, W. (1890). *The principles of Psychology.* New York: Henry Holt & Co.

Jelinek, E. (1995).Quantification in Straits Salish. In: Bach, E., Jelinek, E., Kratzer, A. & Partee, B. (Eds.). *Quantification in Natural Languages.* Kluwer.

Kaplan, D. (1989). Demonstratives. In: *Themes from Kaplan*. Oxford: Oxford University Press.

Köhler S, Paus T, Buckner RL, Milner B. (2004). Effect of left inferior prefrontal stimulation on episodic memory formation: a two-stage fMRI-rTMS study. *J Cognit Neurosci* 16: 178-188.

Klein, S. B., Loftus, J., & Kihlstrom, J. F. (2002). Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition*, *20*, 353–379.

Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology and Aging, 14*, 220–237.

Kurtzman, H.S. (1983). Modern Conceptions of Memory. *Philosophy and Phenomenological Research.* 44(1):1-19.

Laird, J. (1920), *A Study in Realism*, Cambridge: Cambridge University Press.

Larson, R.K. & Segal, G. (1995). *Knowledge and Meaning.* Cambridge: MIT Press.

Larson, R.K. & Ludlow, P. (1993). Interpreted logical forms. *Synthese*. 95(3).

Lashley, K. (1950). In search of the engram. *Society of Experimental Biology* Symposium 4: 454–482.

Lindsay et al. (2004). True photographs and false memories. *Psychological Science*, 15: 149–154.

Lipton, P. (1991), *Inference to the Best Explanation*. New York: Routledge.

Locke, Don (1971). *Memory*. London: Macmillan.

Locke, J. (1979). *An Essay concerning Human Understanding*. NY: Oxford University Press.

Loftus, E.F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7, 560-572.

Loftus, E.F., Miller, D.G., & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 19-31.

Loftus, E.F. & Pickrell, J.E. (1995). The formation of false memories. *Psychiatric Annals*, 25, 720-725.

Lycan, W.G. (1988). *Judgment and Justification*. Cambridge: Cambridge University Press.

Machamer, P.K., Darden, L., & Craver, C. (2000). Thinking about Mechanisms. *Philosophy of Science*. 67(1): 1-25.

Malcolm, N. (1963). *Knowledge and Certainty*. Ithaca: Cornell University Press.

Malcolm, N. (1977) *Memory and Mind*, Cornell University Press: Ithaca, NY.

Martin, C.B. & Deutscher, M. (1966), 'Remembering'. *Philosophical Review*, 75: 161–196.

Martin, M.G.F. (1992). Perception, Concepts and Memory. *Phil Review*. 101(4): 745-63.

Matthews, R. (1989). The Alleged Evidence for Representationalism. In Stuart Silvers (ed.), *Rerepresentation*. Kluwer.

Matthews, R.J. (2007). *The Measure of Mind: Propositional Attitudes and their Attribution.* Oxford: Oxford University Press.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457

McClelland, J.L. (1995). Constructive Memory and Memory Distortions: A Parallel-Distributed Processing Approach. In: D.L. Schacter ed. *Memory Distortion*. Cambridge, M.A.: Harvard University Press.

McClelland, J. L. (2011). Memory as a constructive process: The parallel-distributed processing approach. In S. Nalbantian, P. Matthews, and J. L. McClelland (Eds.), *The Memory Process: Neuroscientific and Humanistic Perspectives.* Cambridge, MA: MIT Press, pp. 129-151

McCormack, T. & Hoerl, C. (1999). Memory and Temporal Perspective: the role of temporal frameworks in memory development. *Developmental Review*. 19: 154–182.

McDowell, J. (1994). *Mind and World.* Cambridge: Harvard Univesity Press.

McKenzie, T. L. B., Bird, L. R. & Roberts, W. A. (2005) The effects of cache modification on food caching and retrieval behavior by rats. *Learning and Motivation* 36(2):260–278.

Melo, B., Winocur, G., & Moscovitch, M. (1999). False recall and false recognition: An examination of the effects of selective and combined lesions to the medial temporal lobe/diencephalon and frontal lobe structures. *Cognitive Neuropsychology,* 16, 343–359.

Michaels, C. F. & Carello, C. (1981). Direct Perception, Englewood Cliffs, NJ: Prentice-Hall.

Mill, J. (1829). *Analysis of the Phenomena of the Human Mind*. London: Baldwin & Cradock.

Millikan, R.G. (1984). *Language, Thought and Other Biological Categories.* Cambridge M.A.: The MIT Press.

Moore, C., & Dunham, P. J. (1995). *Joint attention: its origins and role in development.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Moltmann, F. (2003). Propositional Attitudes Without Propositions. *Synthese*. 135 (1): 77-118.

Moscovitch, M., Rosenbaum, R.S., Gilboa, A., Addis, D.R., Westmacott, R., Grady, C., McAndrews, M.P., Levine, B., Black, S.E., Winocur, G. & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: A unified account based on multiple trace theory. *Journal of Anatomy*, 207, 35-66.

Munsat, S. (1966). *The Concept of Memory*. Random House: NY.

Nadel, L. & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology, 7,* 217-227.

Nadel, L & Moscovitch, M. (1998). The Hippocampal Complex and Long-Term Memory Revisited. *Hippocampus*. 8: 647-650,

Neisser, U. (1967). *Cognitive Psychology*. New York, N.Y.: Appleton.

Neter, J., & Waksberg, J. (1964). A study of response errors in expenditures data from household interviews. *American Statistical Association Journal, 59*, 18 –55.

Nigro, G. & Neisser, U. (1983). Point of View in Personal Memories. *Cognitive Psychology 15*, 467-482

Noë, A. (2004) *Action in Perception*. Cambridge, Mas.: The MIT Press.

Nyberg, L., & Cabeza, R. (2000). Brain imaging of memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory*. New York: Oxford University Press.

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., et al. (2003). Thinking of the future and the past: The roles of the frontal pole and the medial temporal lobes. *Neuroimage*, *19*, 1369–1380.

Paller, K. & Wagner, D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Science.* 6(2): 93-102.

Paller, K. & Voss, J. (2004). Memory reactivation and consolidation during sleep. Learning & Memory, 11, 664-670.

Patterson T. A., Lipton J. R., Bennett E. L., and Rosenzweig M. R. (1990) Cholinergic receptor antagonists impair formation of intermediate-term memory in the chick. *Behav. Neural Biol.* **54,** 63–74.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory & Language, 35*, 261-285.

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple Dissociation of Faces, Bodies, and Objects in Extrastriate Cortex. *Current Biology*, 19(4), 319-324.

Plato (1992). *Theatetus*. Hackett.

Prinz, J. (2007). Mental Pointing: Phenomenal Knowledge without Concepts. *Journal of Consciousness Studies.* 14(9-10): 184-211.

Quine, W.V.O. (1948). On what there is. *Review of Metaphysics.*

Quine, W.V.O. (1968). *Ontological Relativity and Other Essays*. Columbia University Press: NY.

Reid, T. (1785/1849). *Essays on the Intellectual Powers of Man*. Edinburgh: McLachlan, Stewart, & Co.

Reiner, R. & Pierson, R. (1995). Hacking's Experimental Realism: An Untenable Middle Ground. *Philosophy of Science*. 62: 60-9.

Resnik, D.B. (1994). Hacking's Experimental Realism. *Canadian Journal of Philosophy* 24: 395-411.

Richard, M. (1990). *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. New York: Cambridge Univesity Press.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 803-814.

Roediger, H.L. (1996). Memory illusions. *Journal of Memory and Language, 35*, 76-100.
Rosen, D. (1975), 'An Argument for the Logical Notion of a Memory Trace', *Philosophy of Science*, 42: 1–10.

Rosenbaum, S., Koeler, S., Schacter, D.L., Moscovitch, M., Westmacott, R., Black, S., Gao. F., & Tulving, E. (2005). The Case of K.C.: Contributions of a memory-impaired person to memory theory. *Neuropsychologia*. 43(7): 989-1021.

Rubin, D. C., & Greenberg, D. L. (1998). Visual memory deficit amnesia: A distinct amnesic presentation and etiology. *Proceedings of the National Academy of Sciences*. 95, 5413-5416.

Rubin, D.C., Bernsten, D., & Bohni, M.K. (2008). A memory-based model of posttraumatic stress disorder: Evaluating basic assumptions underlying the PTSD diagnosis. *Psychological Review*. 115(4): 985-1011.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Russell, B. (1912). The Problems of Philosophy. *Mind,* 21(84), 116-117

Russell, B. (1913/1992). Theory of Knowledge. Routledge.

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

Ryle, G. (1949). *The Concept of Mind*. Oxford: Oxford University Press.

Rugg, M.D., Johnson, J.D, Park, H., & Uncapher, M.R. (2008). Encoding-retrieval overlap in human episodic memory: A functional neuroimaging perspective. *Progress in Brain Research*, 169, 339-352.

Rugg, M.D., & Wilding E.L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences*, 4(3), 108-115.

Rugg, M. D. & Henson, R.N.A. (2002). Episodic memory retrieval: an (event-related) functional neuroimaging perspective. In A. Parker, E. Wilding and T. Bussey (Eds.) *The cognitive neuroscience of memory: encoding and retrieval.* pp. 3-37. Hove: Psychology Press.

Schacter, D. L. (1995). Memory Distortion: history and current status. In Schacter (ed) *Memory Distortion: how minds, brains, and societies reconstruct the past*, Cambridge, MA: Harvard University Press, pp. 1–43.

Schacter, D.L., Verfaille, M., & Pradere, D. (1996). The neuropsychology of memory illusions: false recall and recognition in amnesic patients. *Journal of Memory and Language*, 35, 319–334.

Schacter, D.L., Norman, K.A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, 49, 289-318.

Schacter, D.L. & Addis, D.R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B, 362, 773-786.*

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). The prospective brain: Remembering the past to imagine the future. *Nature Reviews Neuroscience*, *8*, 657–661.

Schiffer, S. (1987). The `Fido'-Fido Theory of Belief. *Philosophical Perspectives* 1:455-480.

Schiffer, S. (1992). Belief Ascription. *Journal of Philosophy* 89 (10):499-521.

Schroeder, T. (2006). Propositional Attitudes. *Philosophy Compass.* 1: 65-73.

Schwitzgebel, E. (2008). The Unreliability of Naive Introspection. *Philosophical Review* 117 (2):245-273.

Scoville, W.B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiatry*. 20:11–21.

Semon (1904/1921). *The Mneme*. London: Allen & Unwin.

Semon, R. (1909). *Mnemic Psychology*. London: George Allen & Unwin.

Shimamura, A.P. (Forthcoming). CoBRA: A New Theory of Episodic Retrieval.

Shoemaker, S. (1972). Memory. In P. Edwards (ed) *Encyclopedia of Philosophy*, New York: Macmillan, vol. V, 265–274.

Skinner, B.F. (1953). *Science and Human Behavior.* New York: Macmillan.

Sorabji, R. (2006), *Aristotle on Memory.* London: Duckworth.

Stout, G.F. (1915). *A Manual of Psychology*. London: University Tutorial Press.

Spinoza, B. (1985). *The Collected Works of Spinoza.* Vol. I. (E.M. Curley, Ed.) Princeton: Princeton UP.

Szpunar, K. K., & McDermott, K. B. (2008). Episodic future thought and its relation to remembering: Evidence from ratings of subjective experience. *Consciousness & Cognition*, *17*, 330–334.

Steyvers, M., Griffiths, T.L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Science*, 10: 327-334.

Suarez, M. (2008). Experimental Realism Reconsidered: How Inference to the Most Likely Cause Might Be Sound. In: Hartman, S.; Hoefer, C., & Bovens, L. (eds.) *Nancy Cartwright's Philosophy of Science.* Routledge: London.

Suddendorf, T. & Corballis, M. C. (2007). The evolution of foresight: what is mental time travel and is it unique to humans? *Behavioral and Brain Sciences, 30*, 299-313.

Sutton, J. (1998). *Philosophy and Memory Traces: Descartes to connectionism*. Cambridge: Cambridge University Press.

Sutton, J. (2010). Observer Perspective and Acentred Memory: some puzzles about point of view in personal memory, *Philosophical Studies*, *148*, 27–37.

Thomson, R.F. (2005). In Search of Memory Traces. *Ann. Rev. Psych*. 56(1): 1-23.

Treisman, A., Kahneman,D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception and Psychophysics,* 33, 527-532.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*, 381–403. New York: Academic Press.

Tulving, E. (1983). *Elements of Episodic Memory.* Oxford: Clarendon Press.

Tulving, E. (2002). Episodic Memory: from mind to brain. *Annual Review of Psychology*, 53: 1–25

Turvey, M.T, & Shaw, R. (1979). The Primacy of Perceiving: An Ecological Reformulation of Perception for Understanding Memory. In: L-G. Nilsson, *Perspectives on Memory Research: Essays in Honor of Uppsala University's 500th Anniversary.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Underwood, B.J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*, 122–129.

van Fraassen (1980). The Scientific Image. Oxford: Clarendon Press.

Vendler, Z. (1972). *Res Cogitans: An Essay in Rational Psychology*. Ithaca, NY: Cornell University Press

Wagner, A. D., Schacter, D.L., Rotte, M., Koutstaal, W., Maril, A., Dale, A., Rosen, B.R., & Buckner, R.L. (1998). Building Memories: Remembering and Forgetting of Verbal Experiences as Predicted by Brain Activity. *Science*. 281: 1188-1191.

Watson, J.B. (1930). *Behaviorism*. Chicago: University of Chicago Press.

Warnock, M. (1987). *Memory*. London: Faber.

Wheeler, M.E., Petersen, S.E., & Buckner, R.L. (2000) Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences of the United States of America*, (97): 11125–11129.

Wheeler, M. E. & Buckner, R. L. (2003). Functional dissociation among components of remembering: control, perceived oldness, and content. *The Journal of Neuroscience*. 23: 3869-3880

Williams, J. M., Ellis, N. C., Tyers, C., Healy, H., Rose, G., & MacLeod, A. K. (1996). The specificity of autobiographical memory and imageability of the future. *Memory and Cognition*, *24*, 116–125.

Wittgenstein, L. (1953) *Philosophical Investigations,* Blackwell: New York.

Wierzbicka, A. (2007). Is "remember" a universal human concept? In: Amberber, M. (ed). *The Language of Memory in a Cross-linguistic perspective.* John Benjamins: Amsterdam.

Woodruff, C.C., Johnson, J.D., Uncapher, M.R., and Rugg, M.D. (2005). Content-specificity of the neural correlates of recollection. *Neuropsychologia* 43, 1022–1032.

Woodward, J. (1997). Explanation, Invariance, and Intervention. *PSA* 1996, 2: 26–41.

Woodward, J. (2000). Explanation and Invariance in the special sciences. *British Journal for the Philosophy of Science*. 51: 197-254.

Woodward, J. 2002. There is No Such Thing as a Ceteris Paribus Law. *Erkenntnis*. 57(3): 303-328.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, J. & Hitchcock, C. (2003). Explanatory Generalizations, Part I: A Counterfactual Account. *Nôus*, 37: 1–24.