Synthese (2014) 191:155–185
DOI 10.1007/s11229-013-0247-7

Is memory for remembering? Recollection as a form of episodic hypothetical thinking

Felipe De Brigard

Received: 16 November 2012 / Accepted: 16 January 2013 / Published online: 5 February 2013

Abstract Misremembering is a systematic and ordinary occurrence in our daily lives. Since it is commonly assumed that the function of memory is to remember the past, misremembering is typically thought to happen because our memory system malfunctions. In this paper I argue that not all cases of misremembering are due to failures in our memory system. In particular, I argue that many ordinary cases of misremembering should not be seen as instances of memory's malfunction, but rather as the normal result of a larger cognitive system that performs a different function, and for which remembering is just one operation. Building upon extant psychological and neuroscientific evidence, 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 remembering is a particular operation of a cognitive system that permits the flexible recombination of different components of encoded 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.

Keywords Memory · Cognitive function · Remembering · Hypothetical thinking · Core brain network · Episodic future thinking · Episodic counterfactual thinking

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

Thomas Hobbes, Leviathan 1.2.

F. De Brigard (✉)
Department of Psychology, Harvard University, William James Hall, 860, Cambridge, MA 02139, USA
e-mail: brigard@wjh.harvard.edu

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—e.g., the angle at which the plane hit the building, the exact way it broke apart, etc. Despite some disagreements among their answers, 66 % of respondents reported remembering the video vividly (Crombag et al. 1996). But 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 shows that people frequently and ordinarily misremember past experiences (Brainerd and Reyna 2005). The pervasiveness of this psychological phenomenon poses at least two important challenges to the philosophy of memory.[1] The first challenge is epistemological. Traditionally, philosophers have assumed that remembering is factive, so that an utterance of 'S remembers that p' can only be true if p is the case (Bernecker 2010). As a result, cases of misremembering are typically dealt with as cases of "seeming to remember", that is, as mental states whose contents, which are not about personal past events that were the case, are nonetheless experienced as memories (Malcolm 1963; Shoemaker 1967; Audi 1998). From this perspective, then, all those individuals who reported having seen footage of the plane crashing were not really remembering: it merely seemed to them as if they were remembering when they were not. Recently, however, growing scientific evidence supporting the overwhelming pervasiveness of misremembering has invited a number of epistemologists to reassess the normative standards by means of which successful recollection is measured. Some, for instance, have suggested that certain cases of misremembering, such as typical cases of memory distortions and omissions, need not be considered false if one assumes a less stringent normative standard for memory's fidelity to the past (Campbell 2006; Schectman 1994). Likewise, recent "generative" views on memory that take into account its reconstructive nature have tried to reconcile its factivity with some degree of alteration of memorial content

[1] It is an curious 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.

(e.g., Matthen 2010; Michaelian 2011a,c). And, finally, some have gone as far as suggesting that, contra the traditional assumption, "to remember" is not a factive verb after all (Hazlett 2010), effectively undercutting its potential as a reason to support memory's factivity.

But in addition to the epistemological challenge, the phenomenon of misremembering poses a less discussed but equally important challenge for the philosophy of psychology and cognitive science. In particular—I will argue—the pervasiveness of misremembering asks us to reassess our understanding of the function of episodic autobiographical memory.[2] Although not always explicitly articulated, the tacit agreement is that cases of misremembering should be treated as mishaps in an otherwise reliable system whose role is to encode and preserve past information for future use (Kurtzman 1983; Michaelian 2011c).[3] The causes of these failures vary. Some may happen because, during retrieval, information that was supposed to be encoded is missing or gets inappropriately reactivated (e.g., McClelland 1995; McClelland et al. 1995), while others may just be the by-product of a cost-effective tradeoff between the organism's current needs and its processing capabilities (e.g., Anderson and Schooler 2000). Some theorists have suggested that memory failures may represent minor harms to otherwise normally beneficial products (McClelland 1995, p. 75), whereas others have argued that memory failures may in fact be advantageous (Schacter et al. 2011). Either way, the consensus seems to be that cases of misremembering—from minor distortions to contrived confabulations—are to be understood as cases of memory's malfunction. Thus, according to this perspective, all the people that claimed to have seen the video

[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 into 'semantic' and 'episodic memory' (Tulving 1983, 1972). 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. 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). There is some disagreement as to whether or not these terms define perfectly equivalent categories, or even if they name psychological kinds at all (Michaelian 2011b). As it will become clear, the view I am advancing here is sympathetic to the claim that *some* forms of memory may not constitute psychological/natural kinds. Right now, however, I will be sidestepping this issue; I will get back to it briefly in Sect. 4. For the time being, what matters is that the sort of memory experience I will be discussing falls within the psychologist's definition of specific episodic autobiographical memory. Examples include memories about particular—as opposed to general (Conway and Pleydell-Pearce 2000)—events of one's childhood, this or that party I went to in college, the moment in which I received my bachelors degree, or the exact instant in which my wife said 'I do' at our wedding.

[3] Kurtzman (1983), for instance, claims that such is the classic view of memory, so that "any distortion of memory can be attributed to an abnormality in functioning" (p. 3). In a similar vein, Michaelian (2011c) suggests that this "intuitively plausible characterization of memory's function", according to which memory's function "is to preserve information acquired in the past, making it available for future use" (p. 400), has been tacitly assumed by many philosophers and epistemologists. It is worth noting that Michaelian thinks that this characterization is, "at best, a crude oversimplification", and he distances himself from endorsing the claim that misremembering is always memory's failure (Michaelian, in press).

of the plane accident were indeed exercising their memories, but they misremembered because their memory systems malfunctioned.

The purpose of the present essay is to take on this second challenge, and to suggest an alternative explanation as to why certain memory distortions occur. In particular, I will argue that many ordinary cases of misremembering *should not* be seen as instances of memory's malfunction, but rather as the normal result of a larger cognitive system that performs a different function, and for which remembering is just one operation. Thus, according to the view put forth here, most people that claimed to have seen the video of the plane accident did in fact misremember, but their memory systems did not malfunction. To that end, in Sect. 2, I review some critical findings from cognitive science and neuroscience of memory suggesting, first, that misremembering is both normal and pervasive, and second, that failing to misremember may be indicative of a pathological rather than a healthy memory system. These results, I suggest, put pressure on the received view of memory's function, and invites us to approach memory's role in our cognitive economy, not in terms of the contents of the mental states it delivers, but rather in terms of memory's contribution to the cognitive organism. I develop this approach in Sect. 3, and I argue that seeing memory as a cognitive system for remembering the past may not be 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, building upon the work of Schacter and colleagues (e.g., Schacter 2001; Schacter and Addis 2007), I claim that remembering is a particular operation of a cognitive system that permits the flexible recombination of different components of encoded traces into representations of possible past events that might or might not have occurred, presumably in the service of constructing mental simulations of possible future events. Finally, in Sect. 4, I show how my account relates to kindred models of episodic memory, how it can accommodate the evidence discussed in the first part, how it allows us to say that many ordinary instances of misremembering are indeed produced by memory, and how it preserves the intuition that many of such cases 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 upon the fact that when we ordinarily exercise our memory, the contents of the resultant mental states appear to us as being about previous events. This fact has led many philosophers to the natural conclusion that the function of memory is to store, retain, and then to reproduce (or make available, or reconstruct) the contents of past experiences at a later time (e.g., Stout 1915; Locke 1971; Lawlor 2006; Michaelian 2011a). 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.[4] Such a content-based approach—which essentially is a functional characterization based on the domain specificity of a cognitive system—is widely assumed in many discussions about the function of cognitive faculties, including memory. Thus, it is not surprising that, in general, most views in the philosophy of memory consider distorted memories as cases in which memory fails either to faithfully store or to bring back to mind a past experience with fidelity—that is, memory fails to perform its function (Kurtzman 1983).[5]

Despite its intuitive appeal, saying that false and distorted memories[6] are a failure of memory may force 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. Consider some typical distortions in our ordinary experiences of autobiographical recollection: involuntary shifts from *field* to*observer* perspectives, the *telescope effect*, and the *boundary extension* error. Involuntary shifts from *field* to*observer* viewpoints in autobiographical memory are a widely experienced and documented memory phenomenon (Nigro and Neisser 1983). Most of our memories

[4] Arguably, the content-based approach can be traced back to Aristotle. In *De Memoria et Reminiscentia*, for instance, Aristotle explicitly endorses the content-based approach to distinguish the role of memory—the "organ of the soul by which animals remember" (*DM* 553b5-10; Sorabji 2006)—from that of perception and expectation. According to Aristotle, what memory does is different from perception and expectation, for memory's content is the past, whereas the content of perception is the present and the content of expectation is the future (*DM* 449b25; Sorabji 2006).

[5] It is worth mentioning that most philosophers of memory endorse some version of representationalism, according to which, when we remember, we entertain a mental representation depicting an event we experienced in the past. Since memory representationalists are also usually representationalists about perception, remembering is typically understood as either the reproduction or the reconstruction of previous perceptual representations. However, although 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 with Thomas Reid, but it has had some partisans since. However, as it has been pointed out, strong versions of direct realism face difficult metaphysical obstacles, while some of its weaker forms collapse with representationalism (see Locke 1971; Warnock 1987). As a result, for the purposes of this paper, I take memory representationalism as the default philosophical view.

[6] In the psychological literature, the terms "false" and "distorted memories" are often conflated. However, as pointed out by one reviewer, there is an important difference between false and distorted memories. Unless one accepts the strict view that true (or genuine) memories are only those in which the remembered content needs to be identical to the content originally experienced, not all distorted memories would count as false memories. Surely, any version of recollection that acknowledges its reconstructive nature must admit certain degree of distortion during the retrieval of veridical memories. To what extent are distorted memories veridical is an important and difficult question, about which philosophers of mind and epistemologists are currently writing (e.g., Campbell 2006; Michaelian 2011a,c; Sutton 2009, 2010). However, the current paper does not directly speak to such question, as it does not deal primarily with the difference between distorted memories that can and cannot be considered true, but rather with the difference between distorted memories that can and cannot be considered the product of a malfunctioning memory system. As a result, unless otherwise indicated, my use of 'false/distorted memory' is to be understood in opposition to 'true/genuine memory', in the sense of being the product of a functioning system, rather than 'true/veridical memory', in the sense of holding the appropriate truth-relation with the world. In fact, as I just mentioned—and as I will try to make clear in Sect. 4—one of the main purposes of this paper is to argue that false memories, in the epistemic sense, are not the same as memories that are produced by a malfunctioning memory system, as it is often assumed in the philosophical literature.

occur from a 'field' perspective, in the sense that we tend to remember events from the point of view from which we experienced them. Memories from an 'observer' perspective, 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 from an observer perspective, 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 memories from an observer perspective 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 and Waksberg 1964), which refers to people's tendency to remember recent events as being more remote that 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 encountered were larger than they actually are (Intraub and Hoffman 1992). What all of these effects have in common is the fact that they present the remembered content in a somewhat distorted way, that is, as a distortion of the content encoded during the original experience.

The phenomenon of false and distorted memories has also been studied in laboratory settings. 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 et al. (1978) discovered that when the interview was administered 20 min 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 et al. (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 the first demonstration of this effect, Garry et al. (1996) had participants come to the lab for a two-session study. In the first session participants were asked to complete a Life Events Inventory (LEI), in which they had to judge how sure they were that certain events had not happened to them before the age of 10 (e.g., got a stuff animal at a carnival). Two weeks later, the same participants were called back, this time to participate in an imagination task. Participants were asked to imagine what would have happened had they experienced certain childhood events. The childhood events participants had to imagine were randomly chosen from the LEI they completed two weeks before. Importantly,

participants did not remember having imagined any of these critical events back then. At the end of the imagination session, the experimenter pretended to panic and told the subjects that she had lost the original LEI form they filled two weeks earlier, and asked them to fill it again. In reality, this second administration of the LEI was the post-test, as it permitted Garry and colleagues to compare the initial ratings with the post-imagination ratings. Their results showed that participants were more confident saying that the events they did not imagine definitely did not happen in their lifetime than they were about the events they just imagined. More strikingly, Goff and Roediger (1998) demonstrated that participants were more likely to misremember having performed an action in the past, if they had imagined performing said action at a previous time relative to participants that did not imagine it before. The imagination inflation effect has been replicated numerous times, and with a variety of innovative paradigms (Brainerd and Reyna 2005).

So far, I have presented evidence suggesting that misremembering is a normal occurrence in healthy individuals. Another piece of evidence in favor of this claim comes from studies examining performance during memory-related tasks in subjects with episodic memory deficits. In a pioneer study, Schacter et al. (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 these individuals showed significantly reduced false recognition of semantically and perceptually associated lures. Even though they were less accurate than controls overall, individuals with amnesia 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 individuals with amensia, showing that the number of pictorial memory intrusions is significantly reduced relative to healthy controls (Koutstaal et al. 1999; for a review, see Schacter et al. 2002). 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 et al. 2003).

Taken together, these—and many similar—studies suggest several conclusions. First, they tell us that misremembering is a common phenomenon, that it occurs frequently in everyday life, and that it is not simply an artifact of laboratory-based experiments (Schacter 2001). Second, they suggest that even though many of our ordinary memory experiences may have been cases of misremembering, entertaining such distorted memories not only did not affect us: we did not even notice whether we were misremembering. I do not mean to say, of course, that misremembering *never* affects us. Surely many times false memories *do* affect us, but this only happens when we encounter information that contradicts what we thought was the case given that we remember it. However, unless current information makes us aware of such inconsistency, we usually go about our lives without questioning the accuracy of our memories—many of which are, probably, inaccurate. Finally, these studies show that individuals with memory-related pathologies tend to misremember less than normal subjects, which 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. Most ordinary cases of misremembering have an air of plausibility to them (Pezdek et al. 1997). Sometimes we misremember events as if we were distant observers, or as if its boundaries were fully present. Nonetheless, as strange as these particular perspectives may appear to us during recollection, they are not altogether implausible. After all, observer memories are always experienced as from the position a possible observer "might have seen" the event (Nigro and Neisser 1983), and those images whose boundaries are extended during recollection are also perceived as they would have, had one adopted a relatively more distant perspective. 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.[7] Furthermore, recent evidence on the imagination inflation effect shows that imagined possible childhood events participants rate as more plausible are more likely to be falsely recognized as having occurred in a subsequent memory test relative to imagined events that received a low plausibility rating (Pezdek et al. 2006).

At this point one may naturally wonder how can the received 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 I want to put forth is that memory may not be a dedicated system for remembering, but rather that remembering is one of the operations performed by a larger cognitive system that plays a different role in our cognitive economy. However, before I get to do that, there is an obvious objection I need to countenance first. The objection draws on a well-known issue pertaining to the notion of function in biological systems. Even though most biological systems tend to perform their functions regularly, there are many systems that malfunction frequently. However, functions are not statistically determined; the function of a system is not 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

[7] Psychologists call this feature *schema-consistency*, meaning that false memories are consistent with schematic forms of the events they falsely portray. I will discuss this issue further in Sect. 3.

the circumstances in which it evolved, it was a highly reliable indicator of predators.[8] 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 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.[9]

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 are not 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

[8] 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.

[9] A reviewer suggested that it may be worth mentioning, as a third kind of case, Millikan's example against statistical accounts of normal functions: sperm. According to Millikan, "the function of a sperm's tail is to propel it to an ovum, but very few sperm find themselves under normal conditions for proper performance of the tail" (Millikan 1993, pp. 161). Likewise—said the reviewer—one could see each individual memory as normally false or distorted, but perfectly accurate only under very specific ideal circumstances. This is an interesting suggestion, but two considerations dissuaded me from including it in the main text. First, the analogy I am pursuing here is between biological and cognitive systems. Comparing individual memories to individual sperm does not quite capture the kind of malfunction I am going after. Second, as (Boorse 2002, pp. 92–93) pointed out, there is an important sense in which statistical regularities can account for specific circumstances. Millikan's example shows that, on average, most sperm do not perform their function *because* the conditions aren't such that they can successfully do it. However, if circumstances were such that every sperm could perform its function, then sperm's Normal function would presumably coincide with its normal, regular function. This shift in circumstances is also clear in some instances involving biological systems, as it is the case with the first example I mention in the main text.

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 is not 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. Indirect evidence in support of this claim comes from a series of recent studies linking participants' propensity toward false recollection with higher scores in a number of problem-solving tasks. For instance, Howe et al. (2010) demonstrated that falsely recalling a lure during a priming session using the DRM paradigm facilitates solving compound remote associative tasks, a well-known insight-based problem-solving measure of creativity whereby three words are presented (e.g., apple, family, house) and participants are required to quickly come up with one word that can link them all three (e.g., tree). This effect has also been shown in children and older adults (Howe et al. 2011). Similarly, Dewhurst et al. (2011) showed that certain measures of convergent thinking—often considered a main component of creativity—predict susceptibility to false memories, suggesting a correlation between misremembering and creative thinking. Thus, assuming that an increase in problem solving abilities, such as those tapped by these insight-based and convergent thinking tasks, confers cognitive organisms like ourselves an advantage in our current environment, then some tendency toward misremembering may prove advantageous rather than detrimental. The question is, why would this be so?

Inklings of an answer come from a series of recent theoretical proposals on the evolution of our memory systems. 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 memory has to rearrange stored information. Similarly, Boyer (2008) hypothesized that our tendency to deploy episodic memories when forecasting helps counter-motivate our impulsiveness, effectively curbing our projections for future reward in a manner consistent with hyperbolic discounting (Ainslie 2001). As a result, memories need not be perfectly faithful renditions of a past experience: "Memory need be only as 'good' as the advantage in decision-making it affords, measured against the cost of its operation" (Boyer 2009, p. 514). Moreover, it may even be possible that our memory systems evolved to produce relatively inaccurate representations because their psychological utility is more beneficial to the organism than its correspondence with reality (Sutton 2009), be it because it proves beneficial during future planning and forecasting, and/or because it provides emotional, personal or social advantages (Alea and Bluck (2003), see McKay and Dennett (2009), for this line of argument regarding misbelief). At any rate, whatever the precise evolutionary explanation may be, notably

all these theoretical proposals suggest that the flexibility demanded by the cognitive tasks in which episodic memory is deployed require relatively imprecise memory representations.

This idea is not altogether new, however. It was first proposed by Bartlett in 1932, and is rapidly gaining popularity in contemporary cognitive science. Cognitive neuroscientists Schacter and Addis (2007) suggest that considering the conditions in 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 a given time, both literal encoding and recall may become burdensome and risky (Bartlett 1932, p. 204; see also Michaelian 2011c). So, in order to respond quickly while saving storage space, our brains opt for a rather schematic way of encoding information.[10] However, 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.[11] 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 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

[10] 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 where 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 (Parker et al. 2006). Multiple studies have been conducted with amnestic 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.

[11] 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.

poor environments, the idea of memory's fidelity as a function of the environment's informational richness loses its footing.[12]

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 mentioned above, the received view on memory's function followed a content-based approach. Since memories are experienced as reproductions or reinstatements of previous experiences, such an approach recommends thinking of memory's function as that of reproducing or reinstating 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 (Noë 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*.

My suggestion, then, is that in order to understand what memory is for we should look at its function in terms of memory's contribution to cognitive organisms such as ourselves. Philosophers typically distinguish two general approaches to understanding functions in biological systems. The first approach is *etiological*. It consists of analyzing the function of a system in terms of its evolutionary history (e.g., Wright 1973; Millikan 1984). The second approach, which we may call the *role function* approach, consists of analyzing what a system is for in terms of its causal contribution to the organism (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 goal here aligns more with the second approach.[13] Specifically, I want to offer an answer as to what memory may be for by considering memory as a cognitive mechanism whose function is determined by

[12] It is worth noting that, in their paper on the adaptive role of memory distortions, Schacter, Guerin and St. Jacques also distance themselves from the trade-off interpretation I just argued against, and suggest instead that memory distortions may reflect essential adaptive processes such as simulating possible future events, creativity and memory updating. As it will become evident, the proposal I offer in this paper is entirely consistent with their view. In fact, it can easily be seen as suggesting that another one of those adaptive processes leading to occasional false memories is counterfactual thinking. Moreover, Schacter and collaborators suggest that it is a mistake to think of all cases of false memory as cases of memory malfunction, as many kinds of memory distortions "reflect the operation of a normal memory system" (Schacter et al. 2011, p. 472). However, they remain mute as to what the function of such a memory system may be if false memories aren't the result of memory's malfunction. I think of the present paper as speaking directly to that issue.

[13] This is not to say that the first approach isn't worth pursuing. On the contrary, the growing literature on the function of memory generally adopts an etiological approach (see, for instance, Atance and O'Neill 2005; Klein et al. 2002a; Nairne and Pandeirada 2008).

the activities with which it contributes to the overall goals of the containing cognitive organism (see Machamer et al. 2000; Craver 2001). 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 2001, p. 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.

One last caveat: explanations of mechanistic role functions, which are thought to be hierarchical, constitute 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), and maybe Churchland (1986), and Dennett 1987): 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. 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 subroutine of a more complex operation—or so I'm about to argue.

3 Remembering what may happen and what could have happened

There are two parts to answering the question on the mechanistic role function of a system: first, one needs to understand how the mechanisms that *compose* such a system work, and then one has to figure out how such a system contributes to the functioning of the larger system within which it is *contained*. In this section I suggest that extant evidence from cognitive psychology and neuroscience can give us a relatively accurate answer to both parts of this question. To that end, I start illustrating, by way of an example, the role played by the mechanisms involved in recollection during an instance of ordinary misremembering. The purpose of this illustration is to make a case for the thesis that the mechanisms underlying some cases of misremembering are the same as those underlying successful remembering. From this illustration I draw three consequences regarding the nature of such mechanisms. Next, I suggest that these consequences make better sense if we consider memory as an operation of a larger system whose function is not to reproduce (reinstate or reconstruct) past experiences, but rather to recombine them in order to entertain what I call *episodic hypothetical thoughts*. To substantiate this hypothesis—i.e., that remembering

is but an operation of a larger system—I refer to recent supporting scientific evidence.

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 calls 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 into long-term memory (Chun and Turk-Browne 2007). 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.[14]

Now suppose that, 30 min 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 superior temporal and inferior parietal cortices (Shimamura 2011; De Brigard 2011, 2010)—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.[15] Memory, therefore, must be the sort of mechanism that cannot only bind together information that is distributed across neural networks, but must also be able to fill the gaps left

[14] "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 De Brigard 2010).

[15] As pointed out by a reviewer, the analogy with the dinosaur's fossils isn't entirely accurate, as fragments of memory traces do not remain unchanged through time as the fossilized remains of dinosaurs do (see footnote 13). I agree. The point, however, is that not all the fragments need to be there beforehand for the reconstructing to take place.

by the missing pieces (McClelland 1995). Importantly though, the process by means of which memory fills those gaps is not haphazard. On the contrary, there is ample reason to believe that the way in which our memories are reconstructed during recollection is constrained not only by encoding conditions, but also by prior knowledge; specifically, schema-consistent information abstracted from previous experiences with categorically similar items. Let me elaborate.

We have known for quite some time now that previously acquired categorical knowledge influences memory retrieval. Indeed, this very insight allowed cognitive scientists to produce a number of computational models capable of fitting extant data on retrieval effects, such as graceful degradation and assignation by omission (e.g., McClelland and Rumelhart 1985), by way of manipulating the strength of the connections between the target item and categorically related items in memory. Moreover, some of these models have been able to fit results from a restricted number of demonstrations on semantic intrusions (McClelland 1995), whereby lures that have stronger baseline semantic associations with the category of the study list—as in the DRM paradigm—tend to be mistakenly retrieved more often than items with weaker semantic links. But what determines the relative strength of the connections? One promising answer is expertise, understood as the relative frequency of exposure to a particular set of items. Take, for instance, a classic study by Posner and Keele (1968) in which it was demonstrated—among other things—that participants were more likely to make schema-consistent false recognitions to meaningless stimuli (i.e., dot patterns) they previously learned that belonged to an artificial category versus the same stimuli where no previous learning took place. This initial observation—viz., that expertise on a particular category may exert greater influence on false recognition—was tellingly confirmed in a recent experiment by Castel et al. (2007). In this study, participants were asked to study two lists of words: one with names of body parts and one with names of animals. After a brief distraction task, participants were presented with a recognition test, which included some names of body parts and of animals that were not in the study lists. Importantly, all animal words used in this experiment were also names from National Football League teams. Half of the participants had extensive knowledge of American football and followed the sport closely, while the other half had neither knowledge nor interest in the sport. Surprisingly, whereas there were no differences in false recognition of body-parts names, football experts misremembered having seen lure animal names more so than non-experts, suggesting that their expertise in the sport influenced their false recognition rate.

My contention is that expertise—again, understood as the relative frequency of exposure to a particular set of items—may actually help explain why certain fragments of information, and not others, fill in the gaps left by the missing pieces during episodic memory retrieval. Support for this claim can be found in recent Bayesian computational models inspired by the classic framework of the Adaptive Control of Thought-Rational or ACT-R. According to its original version, the ACT-R model conceived of remembering as a computationally expensive cognitive operation whose costs are offset by the gain of achieving a successful recollection (Anderson 1990). Anderson and Schooler (1991) (see also Anderson and Schooler 2000), suggested that the probability of retrieving a target memory can be captured using Bayes' rule by combining the likelihood that a particular cue belongs to a certain context

(i.e., *the context factor*) with the prior probability that the target memory will be needed (i.e., *the history factor*). To determine the context factor—i.e., the likelihood that a particular memory is to be found in a determinate context given a cue—the strategy demands giving a baseline probability to each item in the context given any other item. So, for example, the baseline probability of finding a chair in an office context when one is cued with a desk is higher than when one is cued with a file cabinet, for file cabinets—let's assume—have a lower probability of being found along with chairs in offices than desks do. But determining the history factor is trickier, for it would require "[following] people about their daily lives, keeping a complete record of when they use various facts [and] such an objective study of human information is close to impossible" (Anderson and Milson 1989, p. 705). Anderson and Schooler (1991) solution to this impasse was to extract prior probabilities from the statistical distribution in existing databases that—they thought—could capture "coherent slices of the environment". More recently, however, Hemmer and Steyvers (2009a,b) offered a different, more natural tack: to extract the priors from the participant's own expertise with a certain set of items. For instance, in one study Hemmer and Steyvers (2009a) asked participants to remember particular objects in familiar contexts. Before the experiment, though, they extracted the prior probability of remembering a particular item in a context by gathering participant's judgments on the frequency of finding any item in said context during a norming phase. The distribution of the data collected during the norming phase gave the prior probability plotted in the Bayesian model. Importantly, the model was able to predict both hits and false alarms to items the closer they were to the mean of the prior distribution, strongly suggesting that memory retrieval is sensitive to how frequent we think items could be found in certain contexts. A similar result was obtained by Hemmer and Steyvers (2009b) in a study looking at memory for fruit sizes.

However scant, the evidence provided by these studies, and the Bayesian models derived from them, suggests that the (posterior) probability of remembering a particular item given a context is constrained by the likelihood of that item belonging to said context, combined with the (prior) probability of having seeing that item in the relevant context in the past (see also Huttenlocher et al. 1991; Steyvers et al. 2006). According to this perspective, then, remembering consists in the *optimal reconstruction* of a previous experience, whereby 'optimal reconstruction' is understood as a retrieval process probabilistically constrained both by schematic knowledge and the frequency of prior encounters with the target memory.[16] So, to go back to the example of the Datsun failing to stop at the yield sign: the idea is that since you have had many similar experiences in comparable situations (i.e., with comparable cars and comparable

[16] A reviewer found this claim suggestive, and advised me to relate this line of argument to recent Bayesian models in computational neuroscience, such as Friston (2010) and Clark's (in press), in which my 'optimal reconstruction' could be tantamount to their notion of 'optimization', which is essentially understood as the reduction of surprise or prediction error. Although I am entirely sympathetic to this project, I think that extant evidence of its application to episodic memory is practically non-existent, as most of the models produced within this 'predictive coding' framework pertain to perception and motor tasks. I wouldn't be surprised, though, if the data on the Bayesian models just discussed and the possible results coming from applying the predictive coding framework to episodic memory were entirely consistent. However, wedding the two approaches may bring other complications too, as I discuss in De Brigard (2012).

traffic signs) your perceptual system has grown accustomed to receiving very specific kinds of visual information during perceptual events just like this one. This continuous experience forms a schematic representation of relevantly similar situations, strengthening the connection between the representations of the individual items contained in the scene. The resultant neural network of activation is thus attuned to the probability of seeing 'stop' signs more often than 'yield' signs—or, for that matter, than gigantic lollipops (assuming, for the sake of argument, that stop signs are seen more frequently than yield signs, which in turn are seen more frequently than gigantic lollipops, in similar contexts). 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. 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 and 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 police man—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.

As previously mentioned, there are three consequences I would like to draw from this illustration. First, when I say that certain cases of remembering and misremembering are the workings of the same mechanism, I mean to imply that just as there are cases of misremembering that are like cases of veridical remembering, there are also cases of veridical remembering that are like cases of misremembering. Given that recollection is probabilistic in the sense suggested above, successful encoding just means increased probability of recall.[17] Therefore, a successful recall produced by reconstructing the optimal representation of an event given a cue would count as veridical recall even if some sensory details of the original stimulus were not attended or encoded. 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. Likewise when it comes to the limited space for encoding new information. The brain simply cannot store every detail it attends to, so optimal reconstructions are the best and more efficient alternative for assuring successful retrieval. Most of the time what you recall accurately depicts the witnessed event. Sometimes it does not. In both cases, however, the system is doing what it is supposed to do.

[17] 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 8). A similar idea can be found in the works of Semon (1909), Martin and Deutscher (1966), and more recently Tulving (2002).

The second consequence 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 same regions are playing a different role. Similarly, the superior frontal gyrus, the superior parietal cortex and the hippocampal complex, which are all involved in episodic recollection, have shown to be actively engaged in several distinct cognitive tasks (Anderson 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 consequence, is that our memory system employs a probabilistic strategy to recover information. As a result, the mental content experienced during recollection typically coincides with the content attended to during encoding. But often times memory produces an optimal reconstruction that does not quite map onto the attended event, even if it conveys a highly probable arrangement of memorial fragments for the given cue. Now we can see that 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: ordinary memory distortions have been shown to be not only schema consistent, but also about events that have been deemed plausible by the very subjects who experience such mental states. That is, when people misremember events—or mere details thereof—their false or distorted memories have as contents events that even though 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 to operating with mental contents of events that *did happen* in our past, as well as mental contents of possible events that *could have happened*, even if they did not.[18] What accounts for this remarkable fact? Taking into account the three consequences just drawn, the answer I want to put forth is that the schema-consistency and the plausibility of ordinary cases of misremembering can be explained because the very same probabilistic mechanisms underlying episodic autobiographical recollection are redeployed during the operations of a larger cognitive system that supports what I call *episodic hypothetical thinking*, i.e., self-centered mental simulations about possible events that we think may happen or may have happened to ourselves.

Initial support for this claim comes from recent evidence showing striking parallelisms between the cognitive and neural processes required to remember what happened in our past and to think about what may happen in our personal future (i.e., *episodic future thinking*, Szpunar (2010); for a recent review see Schacter et al. in

[18] The modal operator here is to be interpreted epistemically. Our memory system must be sensitive to alternative ways in which we *think* our past could have been, regardless of whether or not such counterfactual metaphysically obtains. Needless to say, I have in mind a naturalized version of this epistemological interpretation: what we think could have happened in the past is constrained by the psychological mechanisms by means of which we think counterfactually (*cfr*. Williamson 2010). This point will be clarified soon.

press). First, a number of neuropsychological studies have shown that deficits in autobiographical episodic recollection 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 1983; Klein et al. 2002b; Hassabis et al. 2007), older adults with Alzheimer's disease (Addis et al. 2009b) and mild cognitive impairment (Gamboz et al. 2010) patients with severe depression (Dickson and Bates 2005; Williams et al. 1996), individuals with autism (Lind and Bowler 2010), and subjects suffering from schizophrenia (D'Argembeau et al. 2008) and post-traumatic stress disorder (Brown et al. in press). Second, several studies have shown parallels in healthy development of episodic autobiographical memory and future thinking in both children (Atance and O'Neill 2001; Suddendorf and Busby 2005) and older adults (Addis et al. 2008, 2010; Gaesser et al. 2011; Spreng and Levine 2006). Third, a growing number of behavioral studies with healthy individuals has shown that when certain phenomenological 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 and van der Linden 2004; Szpunar and 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 and Schacter 2008; Addis et al. 2007). This core brain network involves the hippocampus, the posterior cingulate/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 could happen, not to themselves, but to other people (Hassabis et al. 2007; Okuda et al. 2003; Szpunar 2010). This suggests that not all kinds of imagining rely on the same brain network (a fact I will revisit soon).

To account for the commonalities between the phenomenological and neural features associated with episodic recollection and future thinking, Schacter and Addis (2007) put forth *the constructive episodic simulation hypothesis*. According to this view, our capacity to think about our future and to bring to mind our past share similar neural structures because both rely on many of the same processes. Specifically, according to their view, in remembering an event we reintegrate representational contents from the encoded experience to reconstruct the unified mental simulation we call recollection, and when we engage in episodic future thinking, the same reintegration processes recombine components from past experiences into a novel simulation of what may occur in our future. This way, the constructive episodic simulation hypothesis agrees with Tulving (1985) famous idea that episodic memory is better understood as a system for "mental time travel", a term taken to refer to our psychological ability to mentally travel back in time to relive past experiences and to project ourselves onto the future in order to anticipate what may come. Consequently, assuming that Addis and Schacter's hypothesis is roughly correct, then we have prima facie evidence to the effect that the same mechanisms underlying thinking of what happened in the past are involved when we engage in a specific kind of hypothetical thinking, namely personal or self-centered—as opposed to non-personal or other-centered—mental simulations of what may happen in our future. And since this variety of future hypothetical thinking is a subspecies of what we may call*episodic hypothetical thinking*—i.e., self-referential

mental simulations about what happened, may happen and could have happened to oneself—then it would also constitute partial evidence to the effect that the very same mechanisms engaged during episodic autobiographical recollection are also responsible for our capacity to entertain episodic hypothetical thoughts.

However, as just mentioned, the evidence presented so far only gives partial support to the claim that the same mechanisms underlying episodic recollection are also engaged during episodic hypothetical thinking. After all, episodic hypothetical thoughts need not be constrained to what could happen in the future: they may also be about what could have happened in our past but did not occur—a subspecies of episodic hypothetical thinking dubbed *episodic counterfactual thinking* (to preserve the symmetry with the term 'episodic future thinking', Szpunar 2010; De Brigard and Giovanello 2012). Could it be possible, then, that the same cognitive and neural structures underlying episodic recollection and future thinking could also be responsible for our capacity to entertain episodic counterfactual thoughts? Although there is very little research on the cognitive and neural mechanisms engaged during counterfactual thinking, and most of it pertains to non-episodic counterfactuals, new psychological and neuroscientific evidence provides evidence suggesting an affirmative answer to this question. To begin with, results coming from developmental psychology suggest that children start engaging in very simple forms of counterfactual thinking by the time they are 3-years-old (e.g., German and Nichols 2003), around the same time long-term autobiographical memories begin to consolidate (Bauer 2006), and also around the same time in which episodic future thinking and planning emerges (Atance and O'Neill 2005). In addition, clues suggesting that individuals with amnesia may have difficulty imagining alternative versions to personal past events come from occasional observations during case studies. For instance, Rosenbaum et al. (2009) report that patient K.C.—who due to a traumatic brain injury developed severe amnesia compromising episodic but not semantic memory—was unable to produce detailed descriptions of imagined events that he could have plausibly experienced in the past, as compared with controls. Similarly, Hassabis et al. (2007) asked individuals with amnesia to generate vivid descriptions of plausible events that could happen to them, not only in the future, but at any time. As compared with controls, these patients offered severely impoverished descriptions of imagined plausible events.[19] Relatedly, a recent behavioral study

[19] Interestingly, one of Hassabis et al. (2007) patients (P01) performed on par with controls, suggesting that not all individuals with amnesia show a concomitant compromise in episodic hypothetical thinking. However, upon closer scrutiny, it was shown that P01 had some remnant anterior hippocampal tissue, apparently comprising part of CA3 and the dentate gyrus, the engagement of which—the authors presumed—may have been sufficient to allow P01 to entertain episodic hypothetical thoughts. Mullally et al. (2012) recently confirmed this suspicion by examining P01's brain activity during the construction of episodic hypothetical thoughts. Strikingly, the remnant hippocampal tissue showed strong activations during successful trials in which episodic hypothetical thoughts were generated. This observation further strengthens the hypothesis that anterior regions of the hippocampus, more so than posterior regions, may be critically involved in the binding of episodic memory fragments (Addis and Schacter 2012). This observation is consistent with recent views in neurobiology suggesting that the capacity to form associations between episodic fragments depend upon the continuous production of new-born granule cells in the dentate gyrus (Deng et al. 2010). Moreover, it is also consistent with recent neurodevelopmental evidence showing that although posterior regions of the hippocampus (i.e., subiculum and CA1) tend to develop relatively early in life, anterior regions (i.e., CA 3 and dentate gyrus) tend to develop at around the time in which episodic autobiographical memory begins to settle (Saitoh et al. 2001).

examining similarities and differences between the phenomenology and the amount of episodic and semantic details during episodic past, future and counterfactual thinking, revealed that the three kinds of mental simulation received equivalent scores on a number of measures, suggesting further similarities among their underlying cognitive processes (De Brigard and Giovanello 2012).

But perhaps the most revealing piece of evidence in favor of the claim that the same cognitive and neural structures engaged during episodic recollection and future thinking are also engaged during episodic counterfactual thinking comes from recent neuroimaging studies. Aware of the fact that common activations during episodic past and future thinking can be explained as the mere "recasting" of memories as future events, effectively eliminating the need of positing any sort of recombination of episodic fragments into novel simulations during future thinking, Addis and colleagues tested their constructive episodic simulation hypothesis against this rival "recasting" view using a novel experimental recombination procedure. In this paradigm, a large number of episodic autobiographical memories are collected from participants during an initial session. Specific details from the participants' memories, such as people, places and objects, are then recombined and presented as cues during a scanning session, allowing researchers to manipulate the recombination of episodic details while keeping their recollective nature constant. For present purposes, however, what matters is that in one of the conditions, Addis et al. (2009a) presented participants with randomly recombined episodes from their past and asked them to imagine a possible past event that could have occurred involving such components. Their results showed that this sort of mental simulation also engaged the core brain network, suggesting that thinking about what could have happened in our past recruits the same brain regions underlying episodic recollection and future thinking. Stronger support for this claim comes from a recent study conducted by Van Hoeck et al. (in press), in which participants were asked to freely imagine alternative positive outcomes to negative events they experienced in their lives. Their results showed that these sorts of episodic counterfactual simulations engaged regions in the right parahippocampal gyrus extending into the hippocampus proper, as well as the middle frontal gyrus, the orbitofrontal and the anterior cingulate cortices. Critically, all these regions belong to the core brain network, suggesting that it not only enables episodic recollection and future projection but also episodic counterfactual thinking. A final piece of evidence comes from a recent study by De Brigard and collaborators (De Brigard et al. in press), in which a variation of the recombination paradigm was employed. In this study, participants provided autobiographical memories about specific events involving either positive or negative outcomes. Later on, during a scanning session, participants were asked to imagine what would have happened had those events occurred with a suggested outcome with the opposite valence. Once again, the core brain network was engaged when participants generated counterfactual thoughts. But, more importantly, De Brigard and colleagues showed that the more likely participants thought the episodic counterfactual event could have been, the more similar the pattern of brain activation was to the neural network engaged during episodic autobiographical recollection. Notably, this result lends credence to the hypothesis offered above to the effect that the involvement of the episodic memory in episodic counterfactual thinking is modulated by the subjective probability with which we think past events could have occurred.

Taken together, the evidence just reviewed suggests that the mechanisms underlying our capacity to remember personal past events 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. Furthermore, the reviewed evidence gives us reason to believe that the very cognitive system by means of which we operate with mental contents about personal events that did happen, also allows us to process mental contents about events that we think may happen or that we think could have plausibly happened in our lives. Remembering, therefore, may be best understood as a particular operation of a larger cognitive system that enable us to entertain episodic hypothetical thoughts. Thus, it is a mistake to think of memory as system that is uniquely—or even primarily—dedicated to reproducing the contents of previous experiences. What we normally call remembering, and what many have tended to identify as the function of an independent cognitive system, is in fact a particular operation of larger system that supports episodic hypothetical thoughts.

4 Misremembering what did not happen as remembering what could have happened

As mentioned at the beginning, partisans of the content-based approach to understanding the function of memory tacitly—and sometimes explicitly—assume that cases of misremembering should be treated as cases in which memory malfunctions. However, in the preceding sections I tried to move away from the content-based approach to thinking about memory's function and instead argued in favor of a mechanistic-role approach. Now I want to suggest that thinking of memory's function from a mechanistic-role approach helps us understand why misremembering is so pervasive, without having to thereby endorse the counterintuitive claim that we have a memory system that constantly and systematically malfunctions. In particular, I want to suggest that when we realize that the very same probabilistic mechanisms underlying episodic autobiographical recollection are redeployed for episodic hypothetical thinking, various phenomena associated with ordinary cases of misremembering can be straightforwardly explicated.

According to the view offered here, when we try to remember an event, memory's underlying retrieval mechanisms reconstruct an optimized mental representation from the encoded perceptual information according to probabilistic constraints dictated by previous experiences. 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 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.

Additionally, the model offered here may help to explain the imagination-inflation effect as well as the phenomenon of schema-consistency in false recollection. As mentioned before, experiments using the imagination-inflation manipulation have revealed that people tend to misremember an imagined event as having occurred more so when they think that it could have plausibly happened to them, relative to imagined events participants considered implausible. But this phenomenon becomes less surprising when we realize that the very same probabilistic mechanisms we use when we engage in episodic counterfactual thinking are redeployed when we attempt to remember episodic autobiographical memories. For according to the view I am suggesting here—which could be seen as a natural extension of Schacter and Addis (2007) constructive episodic simulation hypothesis—generating an episodic counterfactual thought is a matter of binding traces of encoded information to form a mental representation of an event that is likely to have happened in the past, presumably in the service of helping us discipline our thoughts about what may happen in the future. And just as it occurs during episodic recollection, this binding is not haphazard: the informational fragments with which episodic counterfactual simulations are created are sampled from the same set of prior experiences episodic memories are reconstructed from. Moreover, it is very likely that when we generate a counterfactual simulation, these very mechanisms reconstruct the most probable arrangement of constitutive episodic fragments given the context, by way of sampling episodic fragments from the same distribution of stored information memory uses when it fills-in an optimal representation given a cue. That way, when we create episodic counterfactual thoughts on the fly—as when we narrowly miss being hit by a car, for example, and we immediately think of what could have happened instead—the very same mechanisms with which we simulate what could have happened are also updating the priors from which episodic memories are reconstructed, presumably to help adjust our future avoidance strategies. Thus, the more likely is that a particular counterfactual thought could have been a memory—i.e., the more plausible a counterfactual simulation is—the more likely is that it could be taken for one, as it occurs with ordinary cases of plausible and schema-consistent inaccurate recollections.

Finally, I believe that the model offered here also helps to explain why individuals with amnesia tend to produce proportionately fewer distortions and false memories than their healthier counterparts. If the above picture of the neuroanatomical underpinnings of the memory mechanisms 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, individuals with amnesia have not 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 (see Rosenbaum et al. (2009), for a similar view).

Let me now conclude with three final thoughts. First, I hope that this paper helps clarify that there is a distinction between false memories, in the epistemic sense of not meeting the appropriate mind-world relation, and memories that are produced by a malfunctioning memory system. It may be possible that many false memories are produced that way, but they need not be. As a result, simply assuming that false memories are the product of memory's malfunction, as the traditional view appears to hold, is a mistake. Of course, in no way I am suggesting that memory *never* malfunctions, or that *all* false memories are produced by a properly functioning memory mechanism. There are obviously cases in which false memories are produced by retrieval systems working inappropriately, such as cases of confabulation due to damage in the prefrontal and/or the orbitofrontal cortex (Hirstein 2005). My point is simply that not all false memories, in the epistemic sense, are produced by a malfunction in memory, and that it is wrong to jump from the epistemic claim that one's memory is false to the psychological claim that one's memory's system malfunctioned.

The second point pertains to the potential effects of the view I am defending here on the traditional taxonomy of memory systems. As mentioned above (see footnote 2), both philosophers and psychologists have drawn relatively equivalent classifications for different memory systems following, in part, a content-based approach. Since the mechanistic-role perspective I defend here clashes with the content-based approach, and suggests that remembering is merely an operation of a complex cognitive system rather than its sole product, it is natural to think that, if correct, such a view would pose a threat to the traditional taxonomy. Although I am sympathetic to the idea of revising the traditional taxonomy (as are others, e.g., Bergson 1908; Michaelian 2011b), I am not sure that the account I've provided here, along with the evidence cited in its support, is sufficient to question the validity of the entire traditional classification of memory systems. After all, the kind of mental phenomenon I discussed here is limited to concrete episodic autobiographical memories, so any possible consequence would only speak to the relevant sub-categories within declarative memory. Additionally, the traditional taxonomy has some advantages when it comes to categorizing memories phenomenologically. The account I offer here may only affect the cognitive side of the taxonomy, while leaving its phenomenological aspect unscathed. Of course, were it shown that the traditional taxonomy somehow fails to adequately capture said phenomenology, then we would have even more reason to jettison it. But doing that now would be premature. Further research is needed to understand the extent to which a view like the one proposed here would affect traditional classifications of memory systems.[20]

Lastly, a final thought. Throughout this paper I have suggested that the evidence on false and distorted memories may make more sense if we understand remembering as

[20] A related difficulty is whether or not 'memory' is the most appropriate term to refer to the reconstructive mechanisms employed during episodic hypothetical thinking. Further research in the philosophy and the cognitive neuroscience of memory may show that a change in nomenclature could be desirable. I am indebted to a reviewer for raising these two issues.

an operation of a larger cognitive system supporting episodic hypothetical thinking, and I presented some results from cognitive psychology and neuroscience in support of such hypothesis. However, it is important to highlight that in attributing such a function (i.e., episodic hypothetical thinking) to this larger system I am merely making an inference to the best explanation. It has the advantage of agreeing with some extant views, such as Schacter and Addis (2007), but also with Buckner and Carroll (2007), according to which the core brain network's function is to enable us to perform cognitive tasks involving self-projection. Moreover, with not too much maneuvering, the view defended here can also be shown to agree with Boyer (2008) claim that mental time travel plays a fundamental role in decision making. Nonetheless, there is a clear possibility that future research could show that alternative functional accounts may provide a better fit for the evidence. For instance, research on the social functions of memory (e.g., Alea and Bluck 2003) may show that what I've called episodic hypothetical thinking may make more sense in the context of understanding the role autobiographical memory plays within a social group. Similarly, research on the development of autobiographical memory narratives (e.g., Habermas 2007; Fivush et al. 2011) may also reveal that episodic hypothetical thinking ultimately plays a more fundamental role in shaping one's autobiographical identity. Whether or not these views would ultimately end up agreeing with the model proposed here, remains to be seen. As with many things in life, when it comes to memory, time will tell.

Acknowledgments Many thanks to the audiences at the *Society for Philosophy and Psychology* at Lewis & Clark College in Portland, Oregon, the *Metro Experimental Research Group* at NYU, the audiences at the departments of Philosophy and Psychology at the UNC, Chapel Hill, and at the department of Psychology at Harvard. I am also grateful to Donna Rose Addis, Dorit Bar-On, Carl Craver, Daniel Dennett, Shamindra Fernando, Jaclyn Ford, Kelly Giovanello, Adrianne Harris, Bryce Huebner, Justin Junge, William Lycan, Ram Neta, Jesse Prinz, Karl Szpunar, Daniel Schacter, Walter Sinnott-Armstrong, and two reviewers for their valuable comments.

References

Addis, D. R., Pan, L., Vu, M., Laiser, N., & Schacter, D. L. (2009a). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia*, *47*, 2222–2238.

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

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., & Schacter, D. L. (2012). The hippocampus and imagining the future: Where do we stand? *Frontiers in Human Neuroscience*, *5*, 173.

Addis, D. R., Musicaro, R., Pan, L., & Schacter, D. L. (2010). Episodic simulation of past and future events in older adults: Evidence from an experimental recombination task. *Psychology and Aging*, *25*, 369–376.

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., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in episodic simulation of future events. *Psychological Science*, *19*, 33–41.

Ainslie, G. (2001). *Breakdown of will*. Cambridge: Cambridge University Press.

Alea, N., & Bluck, S. (2003). Why are you telling me that? A conceptual model of the social function of autobiographical memory. *Memory*, *11*, 165–178.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
Anderson, J. R. & Milson, R. (1989). Human Memory: An Adaptive Perspective. *Psychological Review*, *96*, 703–719.
Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396–408.
Anderson, J. R., & Schooler, L. J. (2000). The adaptive nature of memory. In E. Tulving & F. Craik (Eds.), *The Oxford handbook of memory* (pp. 557–570). Oxford: Oxford University Press.
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.
Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Science*, *12*, 533–539.
Atance, C. M., & O'Neill, D. K. (2005). The emergence of episodic future thinking in humans. *Learning and Motivation*, *36*, 126–144.
Audi, R. (1998). *Epistemology: A contemporary introduction to the theory of knowledge*. London: Routledge.
Balota, D. A., Cortese, M., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K., et al. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer type. *Cognitive Neuropsychology*, *15*, 361–384.
Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
Bauer, P. J. (2006). Constructing a past in infancy: A Neurodevelopmental account. *Trends in Cognitive Science*, *10*, 175–181.
Bergson, H. (1908). *Matter and memory*. New York: Zone Books.
Bernecker, S. (2010). *Memory*. Oxford: Oxford University Press.
Boorse, C. (2002). A rebuttal on functions. In A. Ariew, R. Cummins, & M. Perlman (Eds.), *Functions. New essays in the philosophy of psychology and biology* (pp. 63–112). Oxford: Oxford University Press.
Boyer, P. (2008). Evolutionary economics of mental time-travel? *Trends in Cognitive Sciences*, *12*(6), 219–223.
Boyer, P. (2009). Extending the range of adaptive misbelief: Memory "distortions" as functional features. *Behavioral and Brain Sciences*, *32*(6), 513–4.
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.
Brown, A. D., Root, J. C., Romano, T. A., Chang, L. J., Bryant, R. A., & Hirst, W. (in press). Overgeneralized autobiographical memory and future thinking in combat veterans with posttraumatic stress disorder. *Journal of behavior therapy & experimental psychiatry*.
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.
Campbell, S. (2006). Our faithfulness to the past: Reconstructing memory value. *Philosophical Psychology*, *19*(3), 361–380.
Castel, A. D., McCabe, D. P., Roediger, H. L, I. I. I., & Heitman, J. L. (2007). The dark side of expertise: Domain specific memory errors. *Psychological Science*, *18*, 3–5.
Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*, 177–184.
Churchland, P. S. (1986). *Neurophilosophy: Toward a unified theory of mind/brain*. Cambridge, MA: MIT Press.
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.
Clark, A. (in press). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral Brain Sciences.
Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*(2), 261–288.
Craver, C. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, *68*, 53–74.

Crombag, H. F. M., Wagenaar, W. A., & 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.
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.
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.
De Brigard, F. (2010). *Reconstructing memory*. PhD Dissertation.
De Brigard, F. (2011). The role of attention in conscious recollection. *Frontiers in Psychology*, *3*, 29.
De Brigard, F. (2012). Predictive memory and the surprising gap. *Frontiers in Psychology*, *3*, 420.
De Brigard, F., & Giovanello, K. S. (2012). Influence of outcome valence in the subjective experience of episodic past, future and counterfactual thinking. *Consciousness and Cognition*, *21*(3), 1085–1096.
De Brigard, F., Addis, D. R., Ford, J. H., Schacter, D. L., Giovanello, K. S. (in press). Remembering what could have happened: Neural correlates of episodic counterfactual thinking. Neuropsychologia.
Deng, W., Aimone, J. B., & Gage, F. H. (2010). New neurons and new memories: How does adult hippocampal neurogenesis affect learning and memory? *Nature Review Neuroscience*, *11*, 339–350.
Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
Dewhurst, S. A., Thorley, C., Hammond, E. R., & Ormerod, T. C. (2011). Convergent, but not divergent, thinking predicts susceptibility to associative memory illusions. *Personality and Individual Differences*, *51*(1), 73–76.
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.
Fivush, R., Habermas, T., Waters, T. E. A., & Zaman, W. (2011). The making of autobiographical memory: Intersections of culture, narratives and identity. *International Journal of Psychology*, *46*(5), 321–345.
Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Review Neuroscience*, *11*, 127–138.
Furlong, E. J. (1948). Memory. *Mind*, *57*, 16–44.
Gaesser, B., Sacchetti, D. C., Addis, D. R., & Schacter, D. L. (2011). Characterizing age-related changes in remembering the past and imagining the future. *Psychology and Aging*, *26*, 80–84.
Gamboz, N., De Vito, S., Brandimonte, M.A., Pappalardo, S., Galeone, F., Lavarone, A., & Della Sala, S. (2010). Episodic future thinking in amnesic mild cognitive impairment. *Neuropsychologia*, *48*(7): 2091–2097.
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 and Review*, *3*, 208–214.
German, T., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science*, *6*, 514–523.
Godfrey-Smith, P. (1994). A modern history theory of functions. *Nous*, *28*, 344–362.
Goff, L. M., & Roediger, H. L. (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory and Cognition*, *26*, 20–23.
Habermas, T. (2007). How to tell a life: The development of the cultural concept of biography across the lifespan. *Journal of Cognition and Development*, *8*, 1–31.
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.
Hazlett, A. (2010). The myth of factive verbs. *Philosophy and Phenomenological Research*, *80*(3), 497–522.
Hemmer, P., & Steyvers, M. (2009a). Integrating episodic and semantic information in memory for natural scenes. In *Proceedings 31st annual conference cognitive science society* (pp. 1557–1562).
Hemmer, P., & Steyvers, M. (2009b). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*, 189–202.
Hirstein, W. (2005). *Brain fiction*. Cambridge, MA: MIT Press.
Howe, M. L., Garner, S. R., Charlesworth, M., & Knott, L. (2011). A brighter side to memory illusions: False memories prime children's and adult's insight-based problem solving. *Journal of Experimental Child Psychology*, *108*(2), 383–393.
Howe, M. L., Garner, S. R., Dewhurst, S. A., & Ball, L. J. (2010). Can false memories prime problem solutions? *Cognition*, *117*(2), 176–181.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in establishing spatial location. *Psychological Review*, *98*, 352–376.
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 and Co.
Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002a). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review*, *109*, 306–329.
Klein, S. B., Loftus, J., & Kihlstrom, J. F. (2002b). 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.
Lawlor, K. (2006). Memory. In *The Philosophy of Mind*. Oxford: Oxford University Press.
Lind, S. E., & Bowler, D. M. (2010). Episodic memory and episodic future thinking in adults with autism. *Journal of Abnormal Psychology*, *119*(4): 896–905.
Lindsay, D. S., et al. (2004). True photographs and false memories. *Psychological Science*, *15*, 149–154.
Locke, D. (1971). *Memory*. London: Macmillan.
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.
Martin, C. B., & Deutscher, M. (1966). Remembering. *Philosophical Review*, *75*, 161–196.
Matthen, M. (2010). Is memory preservation? *Philosophical Studies*, *148*(1), 3–14.
McClelland, J. L. (1995). Constructive memory and memory distortions: A parallel-distributed processing approach. In D. L. Schacter (Ed.), *Memory distortion*. Cambridge, MA: Harvard University Press.
McClelland, J. L., McNaughton, B. L., & 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., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.
McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, *32*, 493–561.
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.
Michaelian, K. (2011a). Generative memory. *Philosophical Psychology*, *24*(3), 323–342.
Michaelian, K. (2011b). Is memory a natural kind? *Memory Studies*, *4*(2), 170–189.
Michaelian, K. (2011c). The epistemology of forgetting. *Erkenntnis*, *74*(3), 399–424.
Michaelian, K. (in press). The information effect: Constructive memory, testimony, and epistemic luck. *Synthese*.
Millikan, R. G. (1984). *Language, thought and other biological categories*. Cambridge, MA: The MIT Press.
Millikan, R. G. (1993). *White queen psychology and other essays for Alice*. Cambridge, MA: The MIT Press.
Mullally, S. L., Hassabis, D., & Maguire, E. A. (2012). Scene construction in amnesia: An fMRI study. *Journal of Neuroscience*, *32*(16), 5646–5653.
Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*, 217–227.

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive memory: Remembering with a stone-age brain. *Current Directions in Psychological Science*, *17*, 239–243.
Neisser, U. (1967). *Cognitive Psychology*. New York, NY: 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, MA: The MIT 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.
Parker, E. S., Cahill, L., & McGaugh, J. L. (2006). A case of unusual autobiographical remembering. *Neurocase*, *12*(1), 35–49.
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 and Language*, *35*, 261–285.
Pezdek, K., Blandon-Gitlin, I., & Gabbay, P. (2006). Imagination and memory: Does imagining implausible events lead to false autobiographical memories? *Psychonomic Bulletin and Review*, *13*(5), 764–769.
Pezdek, K., Finger, K., & Hodge, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science*, *8*(6), 437–441.
Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363.
Roediger III, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.
Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., & Moscovitch, M. (2009). Amnesia as an impairment of detail generation and binding: Evidence from personal, fictional, and semantic narratives in K.C. *Neuropsychologia*, *47*, 2181–2187.
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.
Russell, B. (1921). *The analysis of mind*. London: George Allen and Unwin.
Saitoh, O., Karns, C. M., & Courchesne, E. (2001). Development of the hippocampal formation from 2 to 42 years. *Brain*, *124*(7), 1317–1324.
Schacter, D. L. (2001). *The seven sins of memory*. New York, NY: Houghton Mifflin.
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.
Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (in press). The future of memory: Remembering, imagining, and the brain. Neuron.
Schacter, D. L., Guerin, S. A., & Jacques, P. L. S. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*, 467–474.
Schacter, D. L., Verfaellie, M., & Koutstaal, W. (2002). Memory illusions in amnesic patients: Findings and implications. In L. R. Squire & D. L. Schacter (Eds.), *Neuropsychology of memory* (3rd ed.). New York: Guilford Press.
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.
Schectman, M. (1994). The truth about memory. *Philosophical Psychology*, *7*(1), 3–18.
Semon, R. (1909). *Mnemic psychology*. London: George Allen and Unwin.
Shimamura, A. P. (2011). Episodic retrieval and the cortical binding of relational activity. *Cognitive Affective Behavioral Neuroscience*, *11*(3), 277–291.
Shoemaker, S. (1967). *Memory. The encyclopedia of philosophy*. New York: MacMillan Co. Inc.
Sorabji, R. (2006). Aristotle on memory, 2nd edn. Chicago: University of Chicago Press.
Spreng, N., & Levine., B. (2006). The temporal distribution of past and future autobiographical events across the lifespan. *Memory and Cognition*, *34*, 1644–1651.
Steyvers, M., Griffiths, T. L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Science*, *10*, 327–334.
Stout, G. F. (1915). *A manual of psychology*. London: University Tutorial Press.
Suddendorf, T., & Busby, J. (2005). Making decisions with the future in mind: developmental and comparative identification of mental time travel. *Learning and Motivation*, *36*, 110–125.

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. (2009). Adaptive misbeliefs and false memories. *Behavioral and Brain Sciences*, *32*(6), 535–536.
Sutton, J. (2010). Observer perspective and acentered memory: Some puzzles about point of view in personal memory. *Philosophical Studies*, *148*, 27–37.
Szpunar, K. (2010). Episodic future thought: An emerging concept. *Perspectives on Psychological Science*, *5*, 142–162.
Szpunar, K. K., & McDermott, K. B. (2008). Episodic future thought and its relation to remembering: Evidence from ratings of subjective experience. *Consciousness and Cognition*, *17*, 330–334.
Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.
Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.
Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1–12.
Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25.
Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122–129.
Van Hoeck, N., Ma, N., Ampe, L., Baetens, K., et al. (in press). Counterfactual Thinking: a fMRI study on changing the past for a better future. Social Cognitive and Affective Neuroscience.
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.
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.
Williamson, T. (2010). Reclaiming the imagination. *New York Times*. August 15.
Wright, L. (1973). Functions. *Philosophical Review*, *82*, 139–168.