3.1

Memory and the Intentional Stance

Felipe De Brigard

There are many topics one can't help but associate with the Daniel Dennett: consciousness, free-will, evolution, intentionality, religion. But in discussions of memory, his name may not come up as readily. He mentions the role of memory in consciousness (Dennett 1978; 1991) and dreaming (1976), for instance, but only one paper addresses memory directly, *Mining the Past to Construct the Future: Memory and Belief as Forms of Knowledge,* which he co-authored with his former postdoc, psychologist Chris Westbury, and published in a relatively obscure volume on memory and belief. This paper has been cited just 37 times, with less than 10 citations in philosophy venues. This is unfortunate, in my opinion, because Westbury & Dennett (2000; henceforth W&D) delineate a viable and coherent view of episodic memory that has received substantial support during the last decade and a half of scientific research. Or so I argue in the current paper.

I begin by recapitulating W&D's article in order to highlight their three key theses: a *functional* thesis about how memory works; a *computational* thesis about the process of remembering, and a *metaphysical* thesis about the ontological status of memories. In the rest of the paper I argue that these three theses are not only consistent with one another, but also constitute a coherent view on episodic memory and remembering that has received strong support from different research areas. In section 2, I review recent evidence from cognitive psychology, neuroscience and neuropsychology that lends credence to W&D's functional thesis. In section 3, I review recent work in computational psychology and cognitive science that shows how W&D's computational thesis can be modeled, and that their way of thinking about memory's computational underpinnings may be a good fit for extant data. Finally, in section 4, I offer a reading of W&D's metaphysical thesis that helps to pull together this evidence into a coherent and explanatorily powerful view of the nature of episodic memory: a view according to which remembering is best understood from the intentional stance.

1. An opinionated reconstruction of Westbury & Dennett (2000)

The editors of the volume where W&D appears characterize their contribution as "a useful historical overview of the ways in which philosophers have attempted to delineate the boundary conditions of both memory and belief" (Schacter & Scarry 2000, 4). The authors, though, see their aim as that of clarifying "some conceptual confusions that spring from the way in which we use [the terms 'memory' and 'belief'] in informal discourse" (Westbury & Dennett 2000, 12). But I think their paper offers more than this. My objective, however, isn't to summarize their paper, but rather to reconstruct it as I read it—essentially, as an exercise in reverse-engineering memory (Dennett 1994).

Their chapter starts off with the observation that while some events leave long-term traces, others don't. Events that do not leave long-term traces are "inert historical facts": their past occurrence makes no difference now. For a past event to make a difference now it needs to leave a long-term trace that has the potentiality of becoming operational when it is needed. In the case of experienced events, W&D tell us, "Memory in the fundamental sense is the ability to store useful information and to retrieve it in precisely those circumstances and that form which allow it to be useful" (Westbury & Dennett 2000, 13). There is a strong forward-looking flavor to this understanding of memory. For the recording of traces is not haphazard. Somehow memory must

anticipate what is most likely to become useful at a later time. It would be strategically damaging to save facts that are unlikely to make a difference, as they would be a waste of valuable cognitive storage. Memory must be attuned to store not merely what happens, but also that which is likely to be needed in the future.

The hypothesis that our minds are essentially anticipatory devices has an eminent tradition which, according to one recent account (Hohwy 2013), includes Ibn al Haytham (circa. 1030), Kant (1781) and, more explicitly, von Helmholtz (1855). More recently, many authors have bolstered such a view of the mind by supporting the related hypothesis that brains evolved to essentially become anticipatory machines (Llinás 2001; Clark 2016), a hypothesis that W&D seem to wholeheartedly embrace:

> The whole point of brains, of nervous systems and sense organs, is to produce future, to permit organisms to develop, in real time, anticipations of what is likely to happen next, the better to deal with it. The only way—the only nonmagical way—organisms can do this is by prospecting and mining the present for the precious ore of historical facts, the raw materials that are then refined into anticipations of the future (Westbury and Dennett 2000, 12).

This general principle, W&D reminds us, is especially true of the mechanistic operations of our memory system. Memory should not be seen merely as a passive receptor of information about the past but also as an active producer of anticipatory information about the future (see Ingevar 1985, for a related idea). Moreover, these anticipations of the future—W&D conjecture—are constructed out of the same material as our memories of the past. Together, these two considerations amount to what I take to be W&D's functional thesis:

> *Functional thesis: Our memory system not only processes information about past events but also uses this information to construct useful anticipations of possible future events.*[1]

Thinking of memory is these terms leads to an inevitable question: how can such a system know in advance which information will be useful in the future? W&D's suggestive response relates to another eminent view, usually associated with Bartlett (1932), according to which the encoding and the retrieval of a particular event are influenced by previously acquired knowledge. More precisely, Bartlett's idea was that memory involves "an active organization of past reactions" (Bartlett 1932, 201), so that every new experience is encoded and retrieved not only as an individual event, but also as related to knowledge of similar events we have experienced and encoded. These *schemas* relieve memory from encoding every detail of each event as a unique occurrence—a feat that, given

[1] Further textual support for the functional thesis can be found in one of the very few passages where Dennett talks about episodic memory in *Consciousness Explained*. Here, he suggests that episodic recollection may have co-evolved with our capacity for anticipation: "These techniques of representing things to ourselves permit us to be self-governors or executives in ways no other creature approaches. We can work out policies way in advance, thanks to our capacity for hypothetical thinking and scenario-spinning; we can stiffen our own resolve to engage in unpleasant or long-term projects by habits of self-reminding, and by rehearsing the expected benefits and costs of the policies we have adopted. And even more important, this practice of rehearsal creates a memory of the route by which we have arrived at where we are (what psychologists call *episodic memory*), so that we can explain to ourselves, when we find ourselves painted in the corner, just what errors we made. […] The developments of these strategies permitted our ancestors to look farther into the future, and part of what gave them this enhanced power of anticipation was an enhanced power of recollection—being able to look farther back at their own recent operations to *see* where they made their mistakes" (Dennett 1991, 278).

perceptual and attentional limitations, is simply impossible. In turn, when the time comes to retrieve information, non-encoded details are easily reconstructed in accordance with both schematic knowledge and knowledge of the present situation; or, as W&D put it: "What we recall is not what we actually experienced, but rather a reconstruction of what we experienced that is consistent with our current goals and our knowledge of the world" (Westbury & Dennett 2000, 19).

This reconstructive view of memory implies that memories do not lie dormant and immutable after encoding, but are re-constructed, at the time of retrieval, in accordance with the constrains dictated by previously acquired schemas as well as present conditions of recall:

> The apparently stable objects of memory—the representations of the things being recalled—are not retrieved from some Store-house of Ideas where they have been waiting intact, but rather are constructed on the fly by a computational process. As H.R. Maturana (1970) wrote […] "memory as a storage of representations of the environment to be used on different occasions in recall does not exist as a neurophysiological function" (p. 37). What we call recollection can never be more than the most plausible story we come up with (or, perhaps, only a story which is plausible enough) within the context of the constraints imposed by biology and history (Westbury & Dennett 2000, 19).

This last point—that recollection involves the reconstruction of "the most plausible story we come up with"—suggests that the computational process of reconstructive retrieval W&D have in mind operates probabilistically. Moreover, it suggests that the plausibility of the reconstruction is somehow tied to biology as well as individual history:

> The assessment of what constitutes an acceptable reconstruction of the past must be dynamically computed by an organism under the constraints imposed by its built-in biological biases and the history of the interaction of those biases with the environment in which the organism has lived (Westbury & Dennett 2000, 19).

If we take these historical interactions as referring, at the very least, to the recording of statistical regularities of past events during the generation of the aforementioned knowledge structures or schemas, then we get close to what I take to be W&D's computational thesis:

> *Computational thesis: Our memory system is sensitive to the statistical regularities of the information it encodes, which along with the individual's limitations and current goals, constrain retrieval in a probabilistic manner.*

At this point, W&D change gears and talk about how *philosophers of mind* typically talk about belief, which differs from the way ordinary folk use the term "belief" (Westbury & Dennett, 2000 20). According to W&D, when philosophers of mind talk about belief, they often mean some kind of sentence-like informational structure in the brain. However, since such a view is unlikely (Westbury & Dennett 2000, 20-24; Dennett 1978; 1987), W&D propose to understand belief from the intentional stance:

> Instead of continuing the attempt to define a belief as an entity that an organism might have or not have (in the concrete, binary-valued way that a library can have a particular book or a poem can have a particular line), a belief must be defined in terms of the circumstance under which a belief could be justifiably *attributed* to and

> organism. What is meant when it is asserted that an organism has a belief, we propose, is that its behavior can be reliably predicted by ascribing that belief to it—an act of ascription we call taking the intentional stance (Westbury & Dennett 2000, 24)

This perspective has two immediate consequences. First, what makes an attribution of belief true has nothing to do with whether or not there is a particular brain structure encoding the content expressed by the belief. And second, two individuals—or the same individual at two times—can have the same belief even if what goes on in their brain at each time is different (cf., De Brigard 2015).

W&D go on to suggest that these considerations also apply to memories. One can say, of a certain individual, that she's remembering that *p* even if there is no structure in the brain exclusively dedicated to encoding the content *p*. Moreover, seeing memories from the intentional stance allows us say, of two different individuals, that they are remembering the same memory even if what goes on in their brains is different. Finally, it allows us to say that one can remember the same memory today that one remembered yesterday, even if the neural processes engaged at each time are different. Adopting the intentional stance towards memories, therefore, invites us to think of them from the point of view of what I take to be W&D's metaphysical thesis:

> *Metaphysical thesis: memories do not exist as sub-personal level brain structures encoding particular intentional contents, but rather as personal level psychological phenomena only accessible from the intentional stance.*

Perhaps my reconstruction of W&D's article is mostly wishful reading, but I see them as offering more than "conceptual clarifications" of "memory" and "belief". I take them to be making substantive claims about memory, given its role in anticipation, its reconstructive character, and its personal rather than sub-personal nature. The beauty of this reverse-engineering approach is that the resultant theses are now, fifteen years later, poised for empirical scrutiny, and the verdict of the tribunal of experience favors W&Ds' case.

2. The functional thesis

According to the functional thesis, memory must both process information about past events and employ this information to construct useful anticipations of possible future events. This general hypothesis had received scattered empirical support when W&D wrote about it. Korsakoff (1889/1996) had described an amnesic patient who was both incapable of remembering past experiences and unable to imagine new plans. And Talland (1965) had described instances of chronic Korsakoff's syndrome where patients had difficulty imagining possible future events and remembering past personal experiences. This also appears to have been the case with the amnesic patients H.M. (Buckner, 2010), and K.C. In an oft-quoted passage, Tulving (E. T.) asks K. C. (called here N. N.) what he will be doing the following day (Tulving 1985, 4):

> E. T. "Let's try the question again about the future. What will you be doing tomorrow?" (There is a 15- second pause.)
> N. N. smiles faintly, then says, "I don't know."
> E. T. "Do you remember the question?"
> N. N. "About what I'll be doing tomorrow?"

> E. T. "Yes. How would you describe your state of mind when you try to think about it?" (A 5-second pause.)
> N.N. "Blank, I guess"

K.C.'s inability to remember past personal events and to imagine possible future episodes inspired Tulving to argue that our capacities for episodic memory and for projecting ourselves into the future were underwritten by a general capacity for "mental time travel"

Stronger support emerged in the early 2000s, as a number of studies suggested that episodic memory and future thinking share common neural substrates. One of the strongest pieces of neuropsychological evidence came from patient D.B., who displayed a profound deficit in episodic memory but normal performance in other cognitive tasks, including semantic memory (Klein, Loftus & Kihlstrom 2002). When asked to imagine possible personal future events, D.B. drew a blank, leading Klein and colleagues to posit a common mechanism for mental time travel. Around that same time, Atance & O'Neill (2001) reported evidence of common developmental trajectories for episodic memory and future thinking. And Okuda and colleagues (2003) presented PET data showing common engagement of medial-temporal lobe (MTL) structures during episodic memory and future thinking, in contrast with a control task requiring semantic retrieval. Behavioral studies revealed further parallels between episodic memory and future thinking. For instance, D'Argembeau & Van der Linden (2004; 2006) showed similar effects of valence, temporal distance, and emotion regulation for episodic memory and future thinking. Similarly, Spreng & Levine (2006) showed that the temporal distribution of episodic past and future thoughts could be modeled on a logarithmic scale, as it was more common for subjects to generate events closer to the moment of the simulation than more remote events, suggesting common effects of temporal distance for both episodic memory and future thinking.

Finally, in 2007, three papers provided more conclusive evidence in favor of the claim that some common neural mechanisms are required to remember our past and to imagine possible personal future events. In the first, Szpunar, Watson & McDermott (2007) asked subjects to either remember a past personal event, imagine a possible personal future, or imagine a fictitious event of no relevance for them while undergoing fMRI. Their analysis revealed that thinking about a personal past event and a personal future event, but not thinking about a non-personal event, commonly engaged the medial prefrontal cortex (mPFC)—mainly BA10—the posterior cingulate cortex (pCC), MTL and cuneus. In the second study, also employing fMRI, Addis, Wong & Schacter (2007) cued participants to either remember a past personal episode, imagine a possible personal future event, or create a sentence. Consistent with previous results, they found remarkable overlap between episodic memory and future thinking, as compared to sentence creation, in regions known to be associated with episodic recollection (Cabeza & St. Jacques 2007; St. Jacques & De Brigard 2015). Notably, this common engagement occurred in many of the same regions found by Szpunar and collaborators (2007), with the only difference being the finding of greater hippocampal activity during episodic memory and future thinking relative to the control condition, suggesting both that the hippocampus could have been involved in the non-personal simulations in Szpunar et al.'s (2007) study, and that the hippocampus might be indispensable for mental time travel.

The observation that the hippocampus is critical for both episodic memory and future thinking received even stronger support from the third study, where Hassabis and colleagues (2007) asked five patients with hippocampal amnesia, and ten healthy controls, to imagine possible new experiences in response to verbal cues (e.g., "Imagine you're lying on a white sandy beach in a beautiful tropical bay"). Participants' descriptions of their simulations were transcribed and coded to assess how rich, detailed, and spatially coherent they were. The results of this study clearly showed that amnesic patients' descriptions of their simulations were less rich, contained fewer details, and

were less spatially coherent than the descriptions produced by controls. These results were further replicated by Race and colleagues (2011) on different patients, with the additional demonstration that the deficits in the descriptions were independent of the patient's narrative abilities.

Related evidence regarding the role of the hippocampus in mental time travel comes from the animal literature. Since the finding of hippocampal pyramidal cells whose receptive fields are sensitive to spatial locations, researchers have hypothesized that such "place cells" generate the internal maps we rely on for navigation (O'Keefe & Nadel 1978). These place cells also exhibit regular theta oscillations during active navigation. In two pioneer studies, O'Keefe & Recce (1993) and Skaggs and collaborators (1996) observed tight correlations between theta oscillations and the sequential triggering of place cells—a phenomenon now known as "theta phase precession". Foster & Wilson (2007) then demonstrated that theta phase precessions were time-locked to the activity of corresponding place cells—a phenomenon they dubbed "theta sequences"—which essentially demonstrates that place cells not only code information about specific locations but also about the sequence in which such locations appear, given a certain learned path. Remarkably, Foster & Wilson also discovered that once a rat had learned a path, and was placed in the starting position to run the maze anew, a recap of the relevant theta sequence was recorded, albeit it occurred very quickly—less than 100 ms. In other words, the same neurons that were active during the rat's navigation were quickly reactivated, in the same order in which the relevant locations were to appear, indicating anticipatory activity of the path the rat was about to take (see Buckner 2010 for further elaboration). Convergent evidence for anticipatory activity in the hippocampus comes from a related study by Johnson and Reddish (2007) in which rats were trained on a T-based decision maze while cell ensembles in CA3 were recorded. As expected, they found local spatial mappings coding for specific locations in the maze. But then they used a directionally unbiased algorithm to reconstruct the neuronal activity during points in which the rat had to make a directional decision, and found, at that precise moment, a transient replay of the forward spatial mapping, indicative of a future position in the path rather than the actual position.

The number of studies reporting common engagement of neural structures underlying episodic memory and future thinking has risen steeply since 2007 (Schacter et al., 2012; De Brigard & Gessell 2016), consolidating the view that when we remember our past and imagine what may happen in our future we deploy a common core brain network that significantly overlaps with the so-called *default network* (Schacter, Addis & Buckner 2007; Spreng, Mar & Kim 2009; Spreng & Grady 2010). All such evidence supports W&D's functional thesis to the extent that the neural structures associated with episodic memory and episodic anticipation seem to overlap. But W&D's functional thesis requires more than this, as it claims not only common engagement of cognitive processes but also re-deployment of the *same* information used for remembering when constructing future thoughts. Thankfully, evidence in support of this second part of the claim is also available.

Remembering past autobiographical events appears to require the deployment of a large brain network—likely the default network—of which the hippocampus and adjacent MTL areas are just one set of nodes. Of critical importance is also the re-instatement, at the time of retrieval, of the sensory-cortical regions engaged in during encoding. Initial evidence of this cortical reinstatement came from a study by Wheeler, Petersen & Buckner (2000) in which participants were asked to study words that were paired either with pictures or with sounds while undergoing fMRI. While still in the scanner, participants received a memory test in which they were shown a word and were instructed to retrieve the associated picture or sound. Wheeler and collaborators discovered that pretty much the same peak cluster in the auditory cortex activated during the encoding of the word-sound association, came online during retrieval when cued with a word associated with a sound. Likewise, when the association was between a word and a picture, the same peak cluster of occipital activity registered during encoding was reactivated at retrieval. Since then, a number of neuroimaging studies

have generated further support for the hypothesis that the retrieval of episodic memories requires re-activating modality-specific cortical regions activated during encoding (for a recent review, see Danker & Anderson 2010).

A piece of neuropsychological evidence lends further support to the sensory reactivation hypothesis. Rubin & Greenberg (1998) reported the case of 11 subjects with focal lesions in occipital cortex. Individuals with occipital cortex damage are typically described as having no memory impairment. However, since their damage prevents them from visually encoding pictorial information, assessments of their memory are usually limited to verbal-tasks. But Rubin & Greenberg (1998) developed a strategy to assess the perceptual richness of their episodic memories, as compared to healthy controls, during free recall, and their studies indicated that despite being able to retrieve the gist of past experienced events, individuals with occipital damage retrieved episodic memories that were impoverished and devoid of sensory (particularly visual) details (Greenberg et al 2005). Along with the neuroimaging evidence mentioned above, these results give further support to the claim that the same sensory areas engaged during the encoding of a memory are reactivated when it is retrieved.

Still, this does not necessarily mean that the same information processed by an area at encoding is also processed at retrieval. For all we know, the same area may play different roles at each time (Bergeron 2007). But in a recent study, Barron, Dolan & Behrens (2013) trained participants to associate certain symbols with specific goods ("known goods", e.g., avocado, tea) while undergoing fMRI. Next, participants were asked to imagine "novel goods", which they had never encountered, and which consisted in the combination of two known goods (e.g., avocado tea). They hypothesized that if an imagined novel good involved the combination of two known goods, then they would see reduced activity in the voxels associated with the relevant component known goods as opposed to irrelevant known goods[2]. This is exactly what they found, leading Barron and colleagues, in the spirit of W&D's thesis, to conclude that when people imagine a possible future experience they re-deploy the same sensory information from encoded memories.

The evidence from the mental time travel and the sensory reactivation literatures lends credence to W&D's functional claim, as they suggest that the neural and cognitive mechanisms associated with our capacity to remember our personal past are redeployed to use stored information to construct anticipations of possible future events. Only one piece is missing: these anticipations are supposed to be *useful*. Usefulness is a normatively-laden notion, but some measures are likely proxies. Weiler, Suchan & Daum (2010) asked participants to imagine possible events that may occur in their somewhat immediate future and to rate the *perceived plausibility* of such events while undergoing fMRI. Using ratings of perceived plausibility as parametric modulators, they discovered that hippocampal activity co-varied with perceived plausibility. In a similar vein, De Brigard et al (2013) used spatio-temporal partial least square analyses of fMRI data elicited during episodic recollection or episodic counterfactual thinking, and discovered that subjectively implausible counterfactual thoughts were least similar to patterns of brain activation associated with episodic recollection, while likely counterfactual scenarios recruited default network regions to a much greater extent. Previous research has also shown that, during theory of mind tasks, certain hubs of the default network are modulated by considerations of *social relevance* (Mitchell et al 2006; Krienen et al 2010). Using a combination of reported importance, familiarity and personal closeness, De Brigard and collaborators (2015) conducted an fMRI experiment in which participants were

[2] This method capitalizes on a well known neural phenomenon known as "repetition suppression", which refers to the reduction of neural activation for repeated stimuli relative to novel stimuli. In fMRI, the reduction of BOLD signal for repeated relative to non-repeated stimuli have been widely used to study stimulus and/or feature selectivity (Grill-Spector & Malach 2001).

asked to engage in hypothetical thinking about people they perceived unfamiliar and socially irrelevant, people they considered familiar and socially relevant, or about themselves. Default network activity was evident during simulations involving the self and socially relevant others, but not when participants imagined possible scenarios involving people they didn't know and care much about.

Assuming that perceived plausibility and social relevance are, to some extent, proxies of potential usefulness, the evidence indicates that mental simulations that have more chance of being useful tend to engage the default network to a grater extent than simulations that are perceived as less plausible or less socially relevant. In sum, the empirical evidence reviewed in this section lends strong support to W&D's functional thesis, according to which our memory system not only process information about past events but it also employs this information to construct useful anticipations of possible future events.

3. The computational thesis

According to W&D's *computational* thesis, memory must be sensitive to the statistical regularities of the information it encodes, constraining the information searched during retrieval in a probabilistic manner, in accordance with individual limitations and current goals. Thus expressed, the computational thesis shares a strong family resemblance with John Anderson's *Adaptive Control of Thought*, or ACT, framework. According to ACT-R (the "R" is for "Rational"), remembering is a cognitive operation whose adaptiveness is best captured by *rational analysis*. The basic assumption is that there is always some cost associated with retrieving a memory, and that such cost is offset by the gain attained when retrieval is successful. As such, an adaptive memory system would search for a particular memory as long as the probability of recovering it, given current needs, is greater than the costs of its retrieval. The ACT-R model captures this insight in Bayesian terms: let H_i be the hypothesis that a particular memory is needed during a particular context, and let E be the evidence for an element of said context. Then, $P(H_i|E) \approx P(E|H_i)\ P(H_i)$, where $P(E|Hi)$ determines the likelihood ratio that E is the case given H_i (i.e., the *contextual factor*), and $P(H_i)$ gives the prior probability that a particular memory will be needed (i.e., the *history factor*).

Details aside (cf., Anderson & Milson 1989; Schooler & Anderson, 1997), what concerns us here is that the probability of retrieving a memory is determined by a combination of two factors: the contextual factor and the history factor. The contextual factor attempts to capture known constraints imposed by the conditions of retrieval on the probability of successfully recovering a memory. Consider the "encoding specificity principle" (Tulving & Thomson 1973), according to which the probability of successful retrieval is increased if the information and contextual conditions present during encoding are also available during retrieval. Over the years, the encoding specificity principle has been proven in numerous proprioceptive and environmental contexts—such as studying the material underwater or while sitting at a dentist's chair—and with all sorts of materials: pictures, words, sounds, etc. In addition, other features of the context of retrieval may influence the probability of successfully retrieving a target memory, including attention and current interests and goals. After all, even under fully reinstated conditions of recall, if one's mind is totally distracted, or if one's interests or goals lie elsewhere, retrieval will be hampered. Given the multiplicity of elements present in a retrieval context, Anderson & Milson (1989) suggest that it's better to understand the likelihood ratio as representing the context factor as the multiplicative product of all the likelihood ratios for every element of the context given H_i. As a result, certain contextual elements are better cues than others (i.e., representing a larger positive contribution to the overall product), as it is the case with elements reinstated from encoding to retrieval.

While the contextual factor is captured by the likelihood ratio that a certain E is the case given H_i, $P(E|Hi)$, the history factor is captured by the prior probability that a particular memory would be needed at a time, $P(H_i)$. According to ACT-R, and in agreement with W&D's computational hypothesis, this prior probability depends on the individual's history of previous experiences. Originally, Anderson & Milson (1989, 705) noted that determining the history factor could be daunting, if not impossible, as one "would have to follow 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". As an alternative, Anderson & Schooler (1991) suggest that prior probabilities can be extracted from the statistical distribution of existent databases that would capture "coherent slices of the environment." In an environmental database containing 2 years worth of word usage in *New York Times* headlines, they found that the odds of using a particular word in a headline was inversely correlated with its having occurred in a previous headline, with the probability decreasing the more time had passed since its last usage. Importantly, Anderson & Schooler (1991) showed that this model could capture two well-known memory retrieval effects: the recency effect (Murdock 1962)—whereby people who study lists of items remember items presented toward the end better than those presented in the middle—and the word-frequency effect (Gregg 1976)—whereby high-frequency words are remembered better than low-frequency words. Taken together, the context and the history factors indicate that the probability that a certain memory will be needed in a particular context can be predicted from the probability that it has been needed in the recent past in relevantly similar contexts[3].

Despite ACT-R's impressive results, it seems clear that using priors based on statistical distributions of limited "slices of the environment" does not quite capture the sense in which W&D meant that the history *of the individual* guides the probabilistic reconstruction of the retrieved memory. Indeed, their suggestion was to liken the individual's history to the previously acquired knowledge one brings to bear at the time of retrieval, which—they suggest—may be *schematic* in the sense introduced by Bartlett. Research on the effects of schemas on memory has tended to focus on two issues. The first issue concerns the fact that schematic knowledge increases recognition of schema-inconsistent information relative to schema-consistent information (Bower, Black, & Turner 1979; Rojahn & Pettigrew 1992). For instance, in a list of 20 words where 18 are categorically related and 2 are complete outliers, the probability of remembering those two is higher than the probability of recalling any of the category-consistent ones—provided one controls for serial position effects, frequency, etc. The second issue concerns the effect of schema-consistent relative to schema-inconsistent information on false alarms. For example, in a study by Brewer & Treyens (1981), participants were asked to wait in an office for a few minutes while the experimenter came to fetch them. When the experimenter came back, participants were moved to a different room where they received a surprise recognition test asking them to recall whether or not items on a list were present in the office they were just in. The list included both "old" items—i.e., items that were indeed in the office—and "new" items. Critically, some of the new items—the "lures"—corresponded to objects one would typically find in an office, while the other new items were elements one would rarely find there. Brewer & Treyens (1981) found that participants were much more likely to falsely recognize office-consistent lures than office-inconsistent new items, a finding that has been replicated, in

[3] In the past, I suggested that this kind of model dovetails with Andy Clark's hierarchical predictive approach (2013), as the context and history factors can be combined in a hierarchical model that tries to find the most probable memory—i.e., that which minimizes prediction error—for a needed memory given a certain cue. In other words, I suggested that the ACT-R-based models can be read as describing how memory retrieval attempts to minimize prediction error when finding the optimal memory given the costs of its retrieval and the organism's current needs (De Brigard 2012). I do not pursue this line here further, but see Lin (2014).

various forms, over the last few decades (Lampinen, Copeland & Neuschatz 2001). Many variations on this general finding strongly suggest that schematic knowledge increases false alarms to schema-consistent relative to schema-inconsistent items.

From the Bayesian perspective, these results suggest that false alarms tend to occur for items with high prior probabilities, given a certain sample, while memory correctly rejects items with a low prior probability. What is less clear is the effect of schematic knowledge on "hits"—i.e., correctly recalled items. In a beautiful experiment that speaks to this question, Huttenlocher and colleagues (1991) showed participants images, for just one second, of a black dot in a white circle with a little dot in the center, and asked them to reproduce the location of the black dot after only two seconds of retention interval. They discovered that accuracy in reproducing the location of the dot depended upon two reference strategies that participants, unwittingly, brought to bear at test. The first helped them code the location of the dot in relation to an imagined small circumference around it ("the fine-grain level"). The second helped them code the location of the black dot as falling within an angular sector determined by a radial-polar coordinate system—tantamount to a slice on a pizza pie ("the coarse-grain level"). Huttenlocher and colleagues demonstrated that black dots presented univocally within well delimited fine and coarse grain locations were more accurately reproduced than black dots presented in locations that were not clearly delimited for either or both of the reference strategies (e.g., black dots presented in what would have been a boundary between imaginary pizza slices, or farther from both the center and the outer circumference, were there are no clear reference points to locate the fine-grained imaginary circle). Moreover, they found that the degree of error could be calculated with the same strategy. Huttenlocher et al's model suggests that people use prior topological knowledge to make predictions of where a black dot may be presented, and that black dots falling within the predicted location are better retained. Conversely, there is a monotonic decrease in accuracy linearly related to the degree to which the presented black dot deviates from the predicted location.

Inspired by this line of research, Steyvers & Hemmer (2012) conducted a study looking at the effect of prior knowledge of naturalistic scenes on hit and false alarm rates during a memory test. They employed a two-stage method: first, in a norming session, participants were asked to list the objects they would expect to see in a certain scene (e.g., an urban scene) and to list the objects they could see in a picture of that scene; second, a different group of participants were shown ten images from the set used in the norming session, one at a time, for either 2 or 10 seconds, and they were immediately asked to recall as many items as they could. Consistent with Huttenlocher and colleagues' observations, Steyvers & Hemmer discovered that items that were easily named as belonging to a scene-category and more frequently found in a picture of the relevant scene were better remembered than items that were less easily named and/or less frequently mentioned during the norming session. In fact, for the first four or five items freely recalled during testing, whether participants have seen the scene for 2 or 10 seconds did not make any difference. The combined prior probability of finding an item in scene of a certain type—tantamount to the "coarse-grain level—and of finding an item in a particular instantiation of a certain scene type—akin to the "fine-grained level"—was an equally good predictor of hit rates regardless of whether the encoding time was 2 or 10 seconds.

While Steyvers & Hemmer's strategy (i.e., selecting priors from normed responses given by individuals in the same cohort as the experimental subjects [see Hemmer & Stayvers 2009]) represents an advance over ACT-R's "slice of the environment" approach, it still does not capture the sense in which W&D speak of an individual's memory being constrained by *their own* history and previous knowledge. Anderson & Milson's (1989) concern still looms large: acquiring a schema takes time, and it is not easy to track the process of acquiring a schema to further examine it during a memory test in the controlled environment of the experimental laboratory. For that reason, most

experiments aimed at directly assessing how memory performance is affected by differences in acquired schemas involve comparing within-subject performance for different pre-acquired schemas (Graesser & Nakamura 1982; Roediger & McDermott 1995) or between-subject performance for individuals with different schematic expertise (e.g., de Groot 1966; Chase & Simon 1973; Arkes & Freedman 1984). An example of this last strategy is a study by Castel, McCabe, Roediger & Heitman (2007) in which football experts and non-experts were compared in a recognition test of animal names, some of which were also names of football teams. Castel and colleagues found that, if the animal names also happened to be names of football teams, football experts were better at recognizing previously seen animal names relative to non-experts, but also had more false alarms to animal name lures. Attributing these differences to the participant's schematic knowledge of football is reasonable, yet this kind of experimental paradigm does not allow the direct manipulation of schema acquisition to study their effect on memory performance at the individual level.

Recently, however, researchers have started to explore a different strategy, based upon growing evidence that schema-acquisition may be an instance of category learning (Davis et al., 2014; Love 2013; Sakamoto & Love 2004; Sakamoto 2012). In one study, Palmeri & Nosofsky (1995) asked participants to learn to classify 16 geometric stimuli according to a simple rule. Although most stimuli fit the rule, the learning list included exceptions. A subsequent recognition test showed that subjects recognized category-inconsistent items at a higher rate than category-consistent ones. This recognition advantage for rule-inconsistent relative to rule-consistent exemplars parallels the aforementioned recognition advantage for schema-inconsistent versus schema-consistent items. But the parallels do not end here. In a meta-analysis on schema-dependent recognition memory, Rojahn & Pettigrew (1992) report that the recognition advantage for schema-inconsistent information increases as the proportion of schema-inconsistent to schema-consistent items becomes smaller, a result that was later replicated by Sakamoto & Love (2004) in a category-learning task. And since category-learning can occur relatively quickly, researchers interested in the role of schematic knowledge on recognition memory are starting to employ category-learning paradigms to explore the role of schematic knowledge on recognition memory.

Using novel category-learning paradigm, De Brigard, Brady, Ruzic and Schacter (in press) studied the effect of acquiring new schematic knowledge on memory retrieval. To that end, they first trained participants to categorize computer-created flowers which varied among several features (e.g., color of petals, shape of center, etc.) A critical value was randomly determined as the category-inclusion criterion (e.g., red petals), and it was sampled 50% of the time during the learning period. Unbeknownst to the subject, there was another non-learned category determined by another random feature (e.g., yellow center) that was also sampled 50% of the time. Once they learned the category, participants would study a set of flowers, and would then receive a recognition test. De Brigard and colleagues found that having learned a category increases both hit and false alarm rates during recognition relative to items that did not belong to the learned category. However, items from the non-learned category were also better remembered and elicited more false alarms that items that did not belong either to the learned and the non-learned categories. Following category-learning paradigms, this suggests that people are more likely to correctly recall but also to false alarm to items that are more frequently encountered during learning than to those that are less frequently encountered. However, in a follow-up experiment controlling for the frequency of presentation of each feature during learning, De Brigard et al (in press) showed that when all features are equally sampled, the effect on hit rates goes away, but the false alarm effect for items of the learned category remains, suggesting—just as Anderson and the ACT-R model predicted—that the frequency of prior encounters of a relevant item is only one factor affecting retrieval. Contextual factors, such as current goals and attentional allocation, play also a critical role.

This selective review of recent results in computational modeling and cognitive psychology suggests that the idea of treating memory retrieval as a probabilistic process resulting from both the frequency of prior experiences as well as the contextual factors of retrieval, is feasible and explanatorily powerful. Although the picture is still incomplete, I believe these lines of evidence converge on W&D's computational thesis, according to which the process of retrieving a memory is probabilistic, and depends on priors determined by the statistical regularities of the relevant information it encodes as well as the individual's limitations and current goals.

4. The metaphysical thesis

In section 2 I reviewed evidence coming from cognitive psychology, neuroscience and neuropsychology to support W&D's claim that our (episodic) memory system processes information about the past and mines this information to construct useful anticipations of possible future events. In section 3, I reviewed several lines of evidence from computational psychology and cognitive science that support W&D's claim that the statistical regularities of the information we encode, which along with individual's limitations and current goals, constrain the informational space searched during memory retrieval in a probabilistic manner. In this last section, I argue that these results accord with W&D's contention that memories do not exist as sub-personal brain structures encoding particular contents, but as person-level psychological phenomena describable from the intentional stance. Moreover, I argue that the functional, computational and metaphysical theses form a coherent and strong view on the nature of episodic memory.

Recall W&D's remark that for a past event to make a difference now it must leave a long-term trace with the potential to become operational when it is needed. In the spirit of Dennettian reverse-engineering, let's ask about the nature of memory traces, given the evidence we have marshaled in support of the functional and computational theses. One possibility, rooted in contemporary analytic philosophy, is to consider memory traces as sub-personal beliefs encoded in a language of thought (Fodor 1975). Surprisingly, despite four decades of criticism, appeals to the language of thought to explain the nature of memory traces are alive and well in philosophy. For instance, Bernecker (2008; 2010) has argued that memory traces are dispositional explicit beliefs whose intentional content refers to an experienced event. This definition needs some unpacking. First, a dispositional—as opposed to occurrent—belief is a belief one antecedently held but is currently consciously unarticulated (Bernecker 2010, 84). Second, the content of the belief is explicit, as opposed to implicit, if the representation carrying the content of the belief is "actually present in [one's] mind in the right sort of way, for example, as an item in your 'memory box' or as a trace storing a sentence in the language of thought" (Bernecker 2010, 29). Thus, from Bernecker's perspective, S remembers that p if and only if there is a sub-personal explicit belief S once held, the content of which is represented in some stable trace in S's brain, which has the disposition to become occurrent during conscious retrieval.

Unfortunately, much of what we know about the cognitive psychology and neuroscience of episodic memory speaks against this sub-personal belief view of memory traces. First, the sub-personal belief view assumes that memory beliefs perdure, unchanged, across four discrete and independent stages: belief formation, belief encoding, consolidation or "storage in some 'memory box'" (to use Bernecker's expression), and retrieval and conscious articulation. But contrary to the sub-personal belief view, these processes are neither discrete nor independent. To make this point clear, consider an ordinary experience: you are driving your car when all of the sudden the car in front of you fails to stop at an intersection and gets hit by an incoming truck, all while you maneuver to avoid the collision. The whole event takes no more than 10 seconds. A few minutes later a

policeman arrives and asks you for your witness testimony, which requires you to remember the event.

As you may expect, during the experience of the event you would probably allocate attention to only a subset of elements at the scene (e.g., the wheel, your feet), and for many of them, your attention would only be transiently allocated, if at all (e.g., the grass on the curb, the traffic sign the driver missed). A wealth of psychological evidence shows that retention of episodic information depends on it having been attended during encoding, as attention is considered necessary for conscious perception (De Brigard & Prinz 2010), and conscious perception of a stimulus is required for its successful encoding (Craik & Tulving 1975; for a review, De Brigard 2011b). But this does not mean that attention precedes conscious perception, or that perception precedes encoding. In fact, neural evidence suggests that a number of attention-dependent neural processes occur during conscious perception but persist during encoding, even after attention has been shifted away—e.g., neuronal depolarization (Jensen & Lisman 2005), sustained spiking (Fransen et al. 2002; Hasselmo 2007). A great deal of conscious perception and memory encoding occurs in parallel, and attempts at segmenting the processes into sequential stages are artificial, and do not reflect the psychological complexity of these processes.

Alas, this artificiality is assumed in the sub-personal view of memory representations, as it supposes that there is a particular moment in which the content of a belief is formed, and a subsequent moment in which that very content is encoded. Consider the experience of avoiding the car collision. At what point did you finalize the process of forming a belief about the experience of avoiding the car collision? Did you immediately encode it? When? When you stopped thinking about it? What if you didn't stop thinking about it until the police asked for your testimony? Would that mean that you kept forming the belief but hadn't gotten around to encoding it? There are causal processes underlying the perception of the collision-avoidance event in virtue of which the experience is encoded, but the level of description of these casual processes does not correlate directly with the intentional level of description as is assumed by the sub-personal view of memory representations. The idea of a transparent mapping between intentional descriptions and discrete neural processes is psychologically unrealistic.

An even more pressing difficulty pertains to the moment in which encoding ends and consolidation occurs. According to the so-called "standard model" of consolidation, while encoding requires the interaction of the hippocampus and modality specific areas in the neo-cortex, at some point (which varies from hours to days) the hippocampus is no longer necessary for the maintenance of the memory representation, which in turn becomes stable—consolidated—as a stand-alone neural network poised to be retrieved by the interaction of a cue and the PFC (Squire 1984; McClelland et al. 1995). The standard model accounts for three pieces of evidence from individuals with hippocampal damage: their temporally-graded retrograde amnesia (i.e., Ribot's law), their profound anterograde amnesia but preserved short-term memory, and the fact that both semantic and episodic memory are equally affected. In addition, this standard model of consolidation "at the systems level" is thought to dovetail with theories of consolidation at the synaptic level. The prevalent view on synaptic consolidation holds that experiences are encoded as changes in connectivity among the neurons originally involved in processing the perceptual information. From this perspective, learning consists in the activation and reactivation of neural networks whose co-activation strengthens their connection weights until they become highly selective for their proximal stimulus[4]. Details aside, the moral of the standard model is that experiencing an event, such as the

[4] First articulated by Hebb (1949), this view has found support in molecular and genetic neurobiology (e.g., Kandel 1976; Silva et al. 1998) as well as computational neuroscience (McClelland & Goddard 1996). The precise mechanisms of these

avoidance of the car collision, involves the engagement of several regions of modality-specific sensory cortices: the auditory cortex would process auditory information, such as the sound of the cars crashing; the visual cortex would process visual information, such as the colors and shapes you see through the windshield; the lateral temporal cortices would probably help to categorize the perceived objects on the street, and so on (Frankland & Bontempi 2005). During this process, the hippocampus—presumably modulated by the fronto-parietal attentional network—is binding together these cortical areas into a larger hippocampal-neo-cortical network (McClelland et al. 1995). When binding is no longer required, a memory trace is said to be 'consolidated' in the neo-cortex, in the form of a stable neuronal assembly ready to be reactivated by the pre-frontal cortex given the appropriate cue.

However, three recent lines of scientific evidence suggest that the standard model is, at best, inaccurate. First, a meta-analysis of individuals with MTL amnesia found that their retrograde amnesia for detailed autobiographical events extends for decades, sometimes even for their lifetimes, whereas the retrograde amnesia for semantic memory is less extensive, temporally-graded, and differentially compromised depending on whether it involves public events, world facts or vocabulary (Nadel & Moscovitch 1997). Indeed, the degree of retrograde amnesia is directly proportional to the amount of hippocampal damage, and different regions of the medial temporal lobe differentially contribute to the formation of episodic, spatial, and semantic memories (Nadel et al 2000; Moscovitch et al. 2005). These results indicate not only that the hippocampus may actually be required during both retrieval and encoding (Eldridge et al. 2000; Ryan et al. 2008), but that some memory representations may never reach a point in which they are independent of the hippocampus, and thus consolidated in the standard sense.

The second line of evidence speaks against the idea that, once consolidated, memory representations remain stable and unchanged. A number of studies in animal neurobiology have shown that amnesic interventions that are effective during the initial consolidation of an event are also effective when the supposedly consolidated memory is reactivated, suggesting that the act of retrieving a memory renders its content labile and prone to modifications (Nader & Einarsson 2010; Hardt et al. 2010). Indeed, the view that memories can be updated and modified during reactivation, rather than being stable and unchanging, is becoming the received account of why people are likely to misremember a plausible event as having occurred if they previously imagined it (imagination inflation; Garry et al. 1996), to wrongly recognized as experienced information that was misleadingly introduced at the time of retrieval (post-event misinformation; Loftus & Hoffman 1989), and to false alarm to lures that are conceptually or semantically related to studied items (Roediger & McDermott 1995). Many interpret these results as suggesting that retrieval renders memory traces liable to distortion, as information processed online while one is remembering can infiltrate and/or modify the re-encoded memorial content.

Finally, and as mentioned in Section 2, extant neuroscientific evidence suggests that there isn't a dedicated storage unit for episodic memories. At best, "storage" is a metaphoric label for the following dynamic process: first, connections in the cortico-hippocampal neural network that is engaged during the initial experience of the event are strengthened; second, something like an "index" is formed—likely in the hippocampus—indicating which cells of the distributed neuronal assembly engaged during the encoding of the experience need to be re-activated when remembering takes place (Moscovitch et al, 2005); finally, retrieval consists in the reinstatement of the pattern of activation the brain was in during encoding—a reinstatement that, due to the dynamic, labile and changing process of "storage", is never identical but close enough to its original form. Contrary to

---

neural networks are complex, and involve a number of processes such as enzimatic production (Silva et al. 2002), gene regulation (Kida et al. 2002), and the formation of novel dentritic spines (Engert & Bonhoeffer 1999).

the sub-personal belief view of memory representations, the evidence speaks against there being a trace in the brain that, upon encoding, lies dormant, patiently carrying the intentional content of the encoded experience until it is retrieved. Indeed, the very same regions that were engaged during encoding and are later on recruited during retrieval are constantly redeployed in the interim to serve all sorts of roles, rather than just being dedicated to encode one particular experience.

How can we reconcile all this evidence against the sub-personal belief view of memory with W&D's own remark that, for a past event to make a difference now, it needs to leave a long-term trace with the capacity to become useful when needed? The answer, I surmise, hinges on the term "capacity", for what needs to be stored is not the intentional content of a memory per se, but rather the disposition to entertain such intentional content when it is needed. W&D are clear about this point: being able to retain the "ability to encode useful information and to decode it in precisely those circumstances where it can be useful" (Westbury & Dennett 2000, 24) is, indeed, a process dependent on sub-personal mechanisms. But this sense of storing information in the brain should not be conflated with the ordinary, personal level concept of memory (Westbury & Dennett 2000). To remember that p, then, is not to possess a sub-personal memory belief carrying the relevant intentional content from encoding to retrieval, but to exhibit the kind of behavior that is optimally described and predicted—even for the individual herself—by ascribing the memory that p from the intentional stance.

And here is the great payoff from this austere metaphysical view of memory traces: it dovetails nicely with the evidence marshaled in favor of the functional and computational theses. Rather than evolving to store artificially divided temporal slices of experienced events, our predictive brains developed the capacity to instantiate the dispositional property of reinstating the state they were in during encoding at the time of retrieval. The computational savings of this maneuver are enormous: imagine the amount of storage—the size of the "memory box"—needed for your average-size long-term memory. But our brains are also dynamic machines, engaged in all sorts of exchanges with the external and internal environment that lead to constant changes in its neuronal connections. Given these changes, reinstating the precise state the brain was in during encoding at the time of retrieval may actually be impossible. So the next best solution is to maximize the chances of reinstating this initial state by constraining the reactivation probabilistically, in the sense discussed above. Think, for instance, how easily this view can fit the phenomenon of encoding specificity (Craik & Tulving, 1975). If you help the sensory cortices re-instate the situation they were in, during encoding, at the time of retrieval, by stimulating them with the same stimuli and/or in the same proprioceptive and environmental context in which encoding occurred, this immediately increases the probability of getting the rest of the relevant neural network reactivated. And, finally, I suggest that these very same probabilistic constrains help to explain why the brain would mine the past to construct mental simulations of possible future events. If remembering is the process by means of which we reconstruct the most likely past given the evidence available at the time of retrieval, then why not to employ the exact same machinery to generate anticipations of what may come? It is very unlikely that the future would be exactly as the past was, but it is definitively more likely that it would be as the past could have been. The flexibility afforded by the probabilistic reconstruction constrained by prior experiences is not only an optimal strategy to reconstruct the past but also to predict the future.

6. A (personal) conclusion

There is a sense in which the current paper could be seen as an updated reading of perhaps the only paper Dan Dennett has written on episodic memory, as its main topic. After all, I have presented my argument as an attempt to demonstrate how W&D's functional, computational and

metaphysical thesis, not only find strong support from recent scientific developments but also form a coherent and promising way of understanding memory from a philosophical point of view. But there is another sense in which the current paper is not so much an investigation into how Dennett reverse-engineers memory as much as it is an exercise in how I reverse-engineering my *own views* on memory. It is evident there was much more wishful-reading than there was exegesis in this exercise. And yet, even if mine, I think the view offered here is strictly Dennettian. You see, Dan Dennett shares with Gilbert Ryle, his advisor, an enviable and profoundly generous mentoring trait: a way of influencing one's views without imposing, a way of lovingly letting you discover the advantages of seeing things from his perspective without ever belittling your own. Recently, Dennett articulated this experience in an anecdote, included in the foreword for a recent volume on his first book *Content and Consciousness* (Dennett 1969). Talking about the disorienting experience of working on his dissertation—which would later become this first book—under the direction of Ryle, he mentions how he felt that Ryle never "fought back", but instead tended to agree with his good points, pressing him on mere adjustments, here and there. Dennett even confesses to thinking that he hadn't learned much philosophy from Ryle. But, then, he recalls:

> I finished a presentable draft of my dissertation in the minimum time (six terms or 2 years) and submitted it, with scant expectation that it would be accepted on first go. On the eve of submitting it, I came across an early draft of it and compared the final product with its ancestor. To my astonishment, I could see Ryle's influence on every page. How had he done it? Osmosis? Hypnotism? This gave me an early appreciation of the power of indirect methods in philosophy. You seldom talk anybody out of a position by arguing directly with their premises and inferences. Sometimes it is more effective to nudge them sideways with images, examples, and helpful formulations that stick to their habits of thought (Dennett 2015, vii).

Looking back at my own experience with my doctoral dissertation, I can't help but think that I, too, was Ryled by Dennett. Despite being an external member in my dissertation committee, I had the great fortune of being able to share my thoughts with Dan often as I was writing. At no point, though, did he mention the paper he wrote with Westbury. I learned about this paper only after I had sent him the last draft of my dissertation—he mentioned the W&D paper in passing, more interested, I thought, in telling me how the paper came to be than about its contents. And so I put it aside, and only came to read it much later on, even after two of the chapters of my dissertation were already on their way to being published (e.g., De Brigard 2014a, 2014b). In a way, I felt terrible for not having read it before, as I should have, for after I did, I, too, came to see Dennett's influence on every page I wrote. But perhaps that was exactly Dan's intention. Perhaps that was his indirect way of showing me the advantages of adopting a Dennettian view on memory. I hope the current essay helps to begin articulating such a view, and I also hope it allows me to gratefully express how much I've learned from Dan.

7. Acknowledgements

Thanks to Kourken Michaelian, Bryce Huebner, and Gregory Stewart for useful comments on an earlier draft.

8. References

Addis, D. R. & Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the

future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia 45*(7), 1363-1377.

Anderson, J.R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychol. Rev.* 96, 703–719.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.

Anderson, J.R., and Schooler, L. J. (1991). Reflections of the environment in memory. *Psychol. Sci.* 2, 396–408

Anderson, J. R., and Schooler, L. J. (2000). "The adaptive nature of memory," in *The Oxford Handbook of Memory*, eds E. Tulving and F. Craik (Oxford: Oxford University Press), 557–570.

Arkes, H.R., & Freedman, M.R. (1984). A demonstration of the costs and benefits of expertise in recognition memory. *Memory & Cognition*, 12, 84–89.

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences 5*(12), 533-539.

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

Barron, H.C., Dolan, R.J., and Behrens, T.E. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. *Nat Neurosci.* 16(10): 1492-8.

Bergeron, V. (2007) Anatomical and Functional Modularity in Cognitive Science: Shifting the Focus, *Philosophical Psychology* 20:175-195

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

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

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology,* 11, 177–220.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13, 207–230.

Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annual Reviews Psychology*, *61*, 27-48.

Cabeza, R. & St. Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences 11*(5), 219-227.

Castel, A. D., McCabe, D. P., & Roediger, H. L., III., & Heitman, J. L. (2007). The dark side of expertise: Domain specific memory errors. *Psychological Science*, *18*, 3-5.

Chase, W. G., and Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*, pp. 215–281. New York: Academic Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36(3): 181-204

Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104: 268-94.

Danker, JF., and Anderson, J.R. (2010). The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. *Psychol. Bull.* 136(1): 87-102.

Davis, T., Xue, G., Love, B. C., Preston, A. R., & Poldrack, R. A. (2014). Global neural pattern similarity as a common basis for categorization and recognition memory. *Journal of Neuroscience*, 34(22), 7472-7484.

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 and Cognition 13*(4), 844-858.

D'Argembeau, A. & Van der Linden, M. (2006). Individual differences in the phenomenology of mental time travel: The effect of vivid visual imagery and emotion regulation strategies.

Consciousness and Cognition 15(2), 342-350.
De Brigard, F. & Prinz, J. (2010). Attention and Consciousness. *WIREs Interdisciplinary Reviews: Cognitive Science.* 1 (1): 51-59.
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. Commentary on Andy Clark's "Whatever Next? Predictive Brains, Situated Agents and the Future of Cognitive Science". *Frontiers in Psychology.* 3:420.
De Brigard Addis, D., Ford, J.H., Schacter, D.L., & Giovanello, K.S. (2013) Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia.*51(12): 2401-2414.
De Brigard et al. (2014a). Is memory for remembering? Recollection as a form episodic hypothetical thinking. *Synthese.* 191(2): 155-185
De Brigard, F. (2014b). The nature of memory traces. *Philosophy Compass.* 9(6): 402-414.
De Brigard, F. (2015). What was I thinking? Dennett's *Content and Consciousness* and the reality of propositional attitudes. In: Muñoz-Suárez, C.M. & De Brigard, F. (Eds.). *Content and Consciousness Revisited.* N.Y. Springer. pp. 49-71.
De Brigard et al. (2015) Spreng, R.N., Mitchell, J.P., & Schacter, D.L. (2015). Neural activity associated with self, other, and object-based counterfactual thinking. *NeuroImage.* 109: 12-26.
De Brigard, F. & Gessell, B.S. (2016). Time is not of the essence: Understanding the neural correlates of mental time travel. In: Klein, S.B., Michaelian, K., & Szpunar, K.K. (Eds.) *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel.* NY: Oxford University Press.
De Brigard, F., Brady, T.F., Ruzic, L., & Schacter, D.L. (in press). Tracking the emergence of memories: A category-learning paradigm to explore schema-driven recognition. *Memory and Cognition.* doi: 10.3758/s13421-016-0643-6
de Groot, A.D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving* (pp. 19–50). New York: Wiley.
Dennett, D.C. (1969). *Content and Consciousness*. Routledge & Kegan Paul, London
Dennett, D.C. (1978). *Brainstorms*. Cambridge: MIT Press.
Dennett, D.C. (1987). *The Intentional Stance.* MIT Press/A Bradford Book.
Dennett, D.C. (1976). Are Dreams Experiences? *Philosophical Review*, LXXXV, 151-71.
Dennett, D.C. (1991). *Consciousness Explained.* Little, Brown.
Dennett, D.C. (1994). Cognitive Science as Reverse Engineering: Several Meanings of 'Top-Down' and'Bottom-Up'. In *Logic, Methodology and Philosophy of Science* IX, D. Prawitz, B. Skyrms, and D. Westerståhl, eds., Elsevier Science, BV, Amsterdam, North-Holland, pp. 679-689.
Dennett, D.C. (2015). Foreword to *Content and Consciousness Revisited*, eds. C. Muñoz-Suárez and F. De Brigard, Springer, pp. v-x.
Eldridge LL, Knowlton BJ, Furmanski CS, Bookheimer SY, Engel SA (2000) Remembering episodes: a selective role for the hippocampus during retrieval. *Nat Neurosci* 3: 1149-1152
Engert F, & Bonhoeffer T (1999) Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* 399:66-70.
Fodor, J. 1975. *The Language of Thought*. Harvard University Press
Foster, D. J. & Wilson, M. A. (2007). Hippocampal theta sequences. *Hippocampus, 17*(11), 1093-1099.
Frankland, P.W., & Bontempi, B. (2005). The organization of recent and remote memories. Nat. Rev. Neurosci. 6:119 -130
Fransen, E., Alonso, A.A. and Hasselmo, M.E. (2002) Simulations of the role of the muscarinic-activated calcium-sensitive non-specific cation current I(NCM) in entorhinal neuronal activity during delayed matching tasks. J. Neurosci. 22(3):1081-1097

Garry, M., Manning, C.G., Loftus, E. (1996). Imagination Inflation: Imagining a Childhood Event Inflates Confidence that it Occurred. *Psych Bull Rev*. 3(2): 208-214

Graesser, A. C., & Nakamura, G. V. (1982). The impact of a schema on comprehension and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 59–109). New York: Academic Press.

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

Gregg, V. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), Recall and recognition (pp. 183-216). NewYork: Wiley.

Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol (Amst) 107:293-321.

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

Hassabis, D. & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences, 11*(7), 299-306.

Hasselmo, M.E. (2007). Encoding: Models linking neural mechanisms to behavior. In Science of Memory: Concepts. Eds: R. Roediger, Y. Dudai, and S. Fitzpatrick, New York: Oxford University Press.

Hebb, D. O. (1949). *The organization of behavior*. Wiley: NY.

Hemmer, P. & Steyvers, M. (2009). A Bayesian Account of Reconstructive Memory. Topics in Cognitive Science, 1, 189-202.

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, 98, 352-376.

Ingvar, D.H. (1985). "Memory of the future": an essay on the temporal organization of conscious awareness. *Hum Neurobiol*., 4(3): 127-36

Jensen, O. and Lisman, J.E. (2005). Hippocampal sequence-encoding driven by cortical multi-item working memory buffer. *Trends Neurosci*. 28(2): 67-72.

Johnson, A. & Redish, D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *The Journal of Neuroscience*. 27(45): 12176-12189

Kandel, E. R. (1976). Cellular Basis of Behavior, an introduction to behavioral neurobiology. W. H. Freeman and Company.

Kant, I. (1781) Critique of pure reason.

Kida S, Josselyn SA, deOrtiz SP, Kogan JH, Chevere I, Masushige S, Silva AJ (2002) CREB required for the stability of new and reactivated fear memories. Nat Neurosci 5: 348–355.

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

Korsakoff, S.S. (1889/1996). Medico-psychological study of a memory disorder. *Consciousness and Cognition*, *5*, 2–21.

Krienen FM, Tu PC, Buckner RL (2010). Clan mentality: evidence that medial prefrontal cortex responds to close others. Journal of Neuroscience, 30(41)

Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27, 1211-1222.

Lin, Y. T. (2015). Memory for Prediction Error Minimization: From Depersonalization to the Delusion of Non-Existence - A Commentary on Philip Gerrans. In T. Metzinger & J. M. Windt (Eds). Open MIND: 15(C). Frankfurt am Main: MIND Group

Llinás, R.R. (2001). *I of the vortex*. Cambridge, MA: MIT Press.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General, 118,* 100–104.

Love, B. (2013). Categorization. In K.N. Ochsner and S.M. Kosslyn (Eds.) *Oxford Handbook of Cognitive Neuroscience*, 342-358. Oxford Press.

Maturana, H.R. (1970). *Biology of cognition*. BCL Report 9.0. Biological Computer Laboratory. Department of Electrical Engineering, University of Illinois.

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. & Goddard, N. (1996). Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. Hippocampus, 6, 654-665

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50*, 655-663.

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

Muñoz-Suárez, C.M. & De Brigard, F. (Eds.). *Content and Consciousness Revisited*. N.Y. Springer.

Murdock, B.B. (1962) The serial position effect of free recall. *Journal of Experimental Psychology*. Vol 64(5): 482-488.

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

Nadel, L., Samsonovitch, A., Ryan, L. & Moscovitch, M. (2000). Multiple trace theory of human memory: Computational, neuroimaging and neuropsychological results. Hippocampus, 10: 352-368.

Nader, K. & Einarsson, E. Ö. (2010). Memory reconsolidation: An Update. Ann N Y Acad Sci. 1191: 27-41.

O'Keefe, J. & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford University Press.

O'Keefe, J. & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, *3*(3), 317-330.

Okuda, J. et al. (2003). Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *NeuroImage 19*, 1369-1380.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 21, 548–568.

Race, E. & Keane, M. M. & Verfaellie, M. (2011). Medial temporal lobe damage causes deficits in episodic memory and episodic future thinking not attributable to deficits in narrative construction. *Journal of Neuroscience 31*(28), 10262-10269.

Roediger, 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.

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology, 31*, 81–109.

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

Ryan, L., Cox, C., Hayes, S., & Nadel, L. (2008). Hippocampal Activation during Episodic and Semantic Memory Retrieval: Category Production and Category Cued Recall. Neuropsychologia, 46, 2109-2121.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, 133, 534-553.

Sakamoto, Y. (2012). Schematic influences on category learning and recognition memory. N. M. Seel (Ed.). *Encyclopedia of the Sciences of Learning*. Springer.

Schacter, D.L. & Scarry, E. (Eds.) *Memory, brain, and belief* (pp. 11-32). Cambridge, MA: Harvard University Press.

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

Schacter, D. L. et al. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron*, 76(4), 677-694.

Schooler, L. J., & Anderson, J. R. (1997). The role of process in the ra- tional analysis of memory. *Cognitive Psychology*, 32, 219–250.

Silva, A.J., J.H. Kogan, P.W. Frankland, & S. Kida. (1998). CREB and memory. Annual Review of Neuroscience. 21: 127-148

Silva, A.J., C.F. Stevens, S. Tonegawa, & Y. Wang. (2002). Deficient hippocampal long-term potentiation in alpha-calcium-calmodulin kinase II mutant mice. Science. 257: 201-6.

Skaggs, W.W., McNaughton, B.L., Wilson, M.A., Barnes, C.A. (1996). Theta Phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6:149-72.

Spreng, N. R. & Levine, B. (2006). The temporal distribution of past and future autobiographical events across the lifespan. *Memory and Cognition*, 34(8):1644–1651, 2006.

Spreng, R. N. & Mar, R. A. & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience, 21*(3), 489-510.

Spreng, R. N. & Grady, C. L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of Cognitive Neuroscience, 22*(6), 1112-1123.

Squire, L.R. (1984). Neuropsychology of memory. In: P. Marler and H. Terrace (Eds.) The Biology of Learning. Berlin: Springer-Verlag. 667-685.

Steyvers, M. & Hemmer, P. (2012). Reconstruction from Memory in Naturalistic Environments. In Brian H. Ross (Ed), *The Psychology of Learning and Motivation*, (126-144). Elsevier Publishing

St. Jacques, P. & De Brigard, F. (2015). In D. R. Addis, M. Barense, and A. Duarte, (Eds.), *The Wiley Handbook on the Cognitive Neuroscience of Memory*. John Wiley and Sons, 2015.

Szpunar, K. & Watson, J. M. & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(2), 642– 647.

Talland, G.A. (1965). *Deranged memory: A psychonomic study of the amnesic syndrome*. New York: Academic Press.

Tulving, E. and Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psych Rev.* 80(5): 352-373

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1-12.

Weiler, J., A., Suchan, B., Daum, I. (2010) Foreseeing the future: Occurrence probability of imagined future events modulates hippocampal activation. *Hippocampus*, 20: 685-90.

Westbury, C. & Dennett, D.C. C. (2000). Mining the past to construct the future: Memory and belief as forms of knowledge. In D. L. Schacter & E. Scarry (Eds.) *Memory, brain, and belief* (pp. 11-32). Cambridge, MA: Harvard University Press.

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.