5

A Neurobiological Account of False Memories

VINCENT VAN DE VEN, HENRY OTGAAR, AND MARK L. HOWE

Memory is a reconstructive, rather than a reproductive, system (Loftus, 1991; Nader, Schafe, & Ledoux, 2000; Roediger & McDermott, 1995; Schacter & Loftus, 2013; Schacter, Norman, & Koutstaal, 1998). As a result, we can experience considerable difficulty when trying to distinguish between true and false memories. More specifically, as memories can contain fragments of what was originally experienced, along with (schema-driven) elements that were not experienced but are meaningfully associated with those fragments, it is difficult to decide which parts of our recollections are true and which parts are false. This aspect of memory can lead to harmless or amusing results in discussing anecdotal recollections with friends and family. However, in judicial situations such as in court it can have dire consequences such as false allegations (Loftus, 1993; Roediger & McDermott, 2000), especially when memory serves as (the only) evidence (Howe, 2013; Schacter & Loftus, 2013).

Over the last few decades, human functional neuroimaging studies have increasingly contributed to the understanding of the brain's mechanisms underlying false memories (Mitchell & Johnson, 2009; Schacter & Slotnick, 2004). A neuroscientific description of the brain's memory mechanisms is important to obtain a complete understanding about how human memory works and how memory illusions can arise. Neuroimaging research of false memories addresses the question as to whether brain activity of a true memory can be distinguished from that of a false memory. This chapter reviews developments and advances in this field of research, and discusses possibilities and pitfalls in translating neuroscientific findings to the courtroom.

Neuroimaging methods, and particularly functional magnetic resonance imaging (fMRI) and electro-/magnetoencephalography, have been used to map memory-related processes on localized brain areas and structures. A plethora of neuroimaging studies have revealed that memory formation and retrieval activate a distributed brain-wide network of cortical areas and subcortical structures that support associative processes (medial temporal lobe, including hippocampus [Paller & Wagner, 2002; Squire, 1992]), auditory (e.g., Rauschecker &

Scott, 2009) and visual perception (Tootell, Hadjikhani, Mendola, Marrett, & Dale, 1998), attentional selection and enhancement of processing (e.g., frontal and parietal areas [Corbetta & Shulman, 2002; Hopfinger, Buonocore, & Mangun, 2000]), cognitive control and executive functions (frontal areas [Van Veen & Carter, 2002; Weissman, Roberts, Visscher, & Woldorff, 2006]), and social and emotional information processing (amygdala [(Phelps & LeDoux, 2005] and medial ventromedial frontal cortex [Schilbach, Eickhoff, Rotarska-Jagiela, Fink, & Vogeley, 2008]). Figure 5.1 provides an overview of the anatomical locations of many of these brain areas. Many studies have shown that the brain areas that are associated with false memories, for a large part, overlap with those of accurate memory judgments (Buckner & Wheeler, 2001; Schacter & Slotnick, 2004), in line with the notion that true and false memories rely on shared memory mechanisms. However, there are some indications that brain activity for false memories can be distinguished from true memories under certain experimental conditions.

Many excellent reviews about the cognitive neuroscience of false memories have already been published (Buckner & Wheeler, 2001; M. K. Johnson, Raye, Mitchell, & Ankudowich, 2012; Mitchell & Johnson, 2009; Schacter & Slotnick, 2004). This chapter reviews previous findings, but also extends previously published reviews to include more recent developments in cognitive neuroscience, as well as a few examples from molecular neuroscience. First, it discusses relevant findings from human brain imaging research that show the involvement of various cortical areas and brain structures in true and false memories. Second, it discusses what neuroscience studies can and cannot tell us about false memories

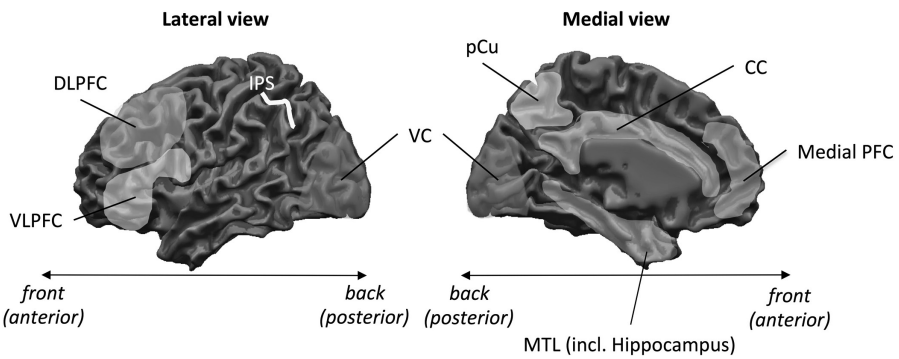

Figure 5.1 Anatomical locations of brain areas and structures related to true and false memories. Shown is a computer-generated representation of the left hemisphere from a lateral (seen outside looking inward) and medial view (seen from in-between the hemispheres outward). The "hills and valleys" (i.e., gyri and sulci) of the cortical curvature are slightly exaggerated. Dark gray areas are valleys (sulci) and lighter gray areas are hills (gyri). Drawn on top of the curvature are representations of the various brain areas and structures. Abbreviations: DLPFC, dorsolateral prefrontal cortex; VLPFC, ventrolateral prefrontal cortex; VC, visual cortex; pCu, precuneus; CC, cingulate cortex; MTL, medial temporal lobe.

in laboratory settings and in real life. Third, it briefly describes efforts in fundamental and molecular neuroscience that demonstrate how principles of synaptic connectivity may be associated with the generation of false memories. The chapter concludes with a few critical considerations about the possible role of neuroimaging in the courtroom.

A NEUROBIOLOGICAL ACCOUNT OF FALSE MEMORIES

False memories are recollections of events that did not happen, or that did not happen in the remembered context (Roediger & McDermott, 1995, 2000). The term *false memory* does not refer to a theoretical framework or model about memory, but rather to the phenomenon of reporting events from memory for which there is no evidence of an external correlate. There is strong consensus that the mechanisms of false memories are largely similar to those of accurate memory formation and retrieval (Mitchell & Johnson, 2000; Roediger & McDermott, 1995; Schacter & Slotnick, 2004). Rather than a separate class of memories, false memories reflect the erroneous decision that a mental experience derives from a recollection of a previously experienced event.

Understanding False Memories as Misassociations

A key notion in understanding memory, both true and false, is that it is associative in nature (Howe, Wimmer, Gagnon, & Plumpton, 2009; Tulving & Craik, 2000). Typically, false memories pertain to episodic or autobiographical memories, which are recollections of perceptual or emotional events that are embedded within a spatial, temporal, or cognitive context in which the events occurred. Furthermore, recollection from episodic or autobiographical memory is considered to be cue dependent, that is, current mental experiences can lead to recollections of previous mental experiences from memory. Cues can be explicit items or impressions (such as the face of a friend recalls fun times) or implicit contextual information (such as doing a memory test in the same room that you studied for the test facilitates your performance). A prominent cognitive framework about the contextual nature of memory is the Source Monitoring Framework (SMF) (M. K. Johnson, Hashtroudi, & Lindsay, 1993; Mitchell & Johnson, 2000). The SMF states that the recollection of an event is an attribution about a current mental experience, which is based on the association between (previous) mental experiences of an event that have been bound together (paraphrased from p. 20 in M. K. Johnson et al., 2012). These mental experiences can include sensory/perceptual experiences, semantic information, contextual information, emotions, and information about cognitive operations, as well as physiological states, which influence the judgment about the source of a current mental experience, as "different sources differ on these dimensions" (M. K. Johnson et al., 2012, p. 20). The notion of *source* can refer to the spatiotemporal origin of an event (the time or place in which something occurred), the acting agent from which the event occurred (events that you

caused yourself or by someone else) or some other context that is relevant to the event (e.g., a semantic "source"). This framework predicts that false memories result from misattributing or erroneously associating a mental experience to an unrelated context. The SMF has a strong theoretical basis because it provides testable predictions about how episodic memories are formed and how this could lead to false memories (M. K. Johnson et al., 2012; Mitchell & Johnson, 2009). In what follows, the text will refer back to the SMF when interpreting the presented fMRI studies.

Mapping the Neural Correlates of Memory

The workings and applications of fMRI are well explained to the nonexpert reader in a number of excellent publications (Amaro & Barker, 2006; Huettel, Song, & McCarthy, 2004; Logothetis, 2008; Poldrack et al., 2008). This chapter skips methodological details and considers only the key conceptual components. The common approach to investigating the neural correlates of true and false memory is to interpret changes in amount of brain activity (or fMRI *signal amplitude*) in relation to some memory function or judgment. Brain areas that show increased activity during moments of recollection compared with moments of rest (in which there is no memory recollection) are interpreted as being associated with memory recollection. At the same time, brain areas that show increased activity during recollection *and* during another mental act could indicate that the respective brain area supports a function that is shared between mental processes. For example, increased activity during recollection in brain areas that are associated with visual perception could indicate perception-related processes during memory recollection, such as the imagining of visual details of the remembered event. Figure 5.2 provides a schematic overview of this approach. Crucial to this approach is the ability with which the investigator can identify the process of interest (e.g., whether a participant is recalling a mental event from memory), but also the ability to identify some control process in which the process of interest is missing (e.g., doing a mathematical assignment). The difference in amount of brain activity as a function of the difference between the targeted and control process can then serve as evidence to whether the brain area is associated with the process of interest. Clearly, choices that affect how the targeted and control processes are manifested will strongly influence the outcome and interpretability of the findings.

In false memory research, investigators are arguably interested in two brain activity scenarios. In the first scenario, the premise is that false memories are not the same as true memories, and the aim is to find brain areas that are more activated during the encoding or retrieval of true memories than for false memories. Identification of such brain areas could elucidate how false memories can be distinguished from true memories on a cognitive or neural level. Importantly, such a true/false memory distinction at the level of brain activity does not mean that the participant is conscious of, or somehow has unconscious access to, this distinction.

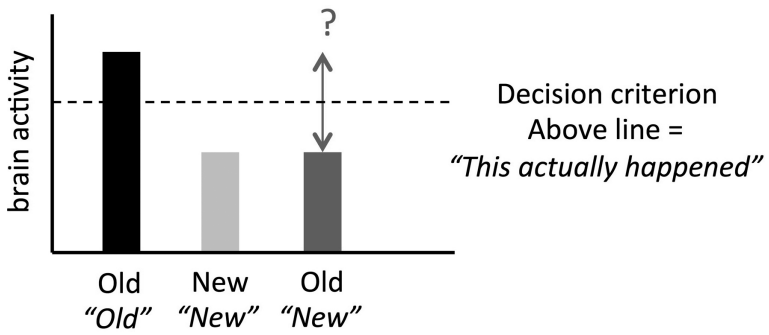

Figure 5.2 Schematic representation of the differential amplitude approach in mapping memory functions on brain areas. The premise is that memory-related areas show increased activity for true memory formation and/or retrieval (Old items that the participant judges as "old," black bar), in comparison with seeing items that were not memorized from a previous experience (New items that the participant judges as "new," gray bar). These brain areas can respond to false memories (New items judged as "old," red bar) in various ways (indicated by red question mark). Brain areas that process true memories different from false memories show less activity for false memories than for true memories. Brain areas that do not dissociate between true and false memories show activity similar to that of true memories. A false recollection can be judged as a true memory if brain activity surpasses a certain threshold (dashed horizontal line) and becomes similar to that of true memories.

In the second scenario, the premise is that false memories are misinterpreted as being true memories because of some shared property between the two memory types. Here, the aim is to find brain areas in which activity for false memories matches the activity for true memories. Thus, if brain activity of a mental event becomes more similar to the pattern of activity that is associated with a true memory, then that mental event is more likely to be interpreted as a true memory.

Neural Correlates of False Recollections

In cognitive psychological research, the Deese-Roediger-McDermott (DRM) paradigm has often been used to elicit false memories in healthy participants (Deese, 1959; Roediger & McDermott, 1995). In the DRM task, participants are presented with lists of words that are related to an unpresented concept (e.g., candy, chocolate, icing, sugar, nice, sour, and cake are related to the unpresented item SWEET). During subsequent recall or recognition testing, participants often report the unpresented item, also termed *the critical lure*, along with items from the previously presented lists.

In a number of early neuroimaging studies, brain activity was measured while participants completed a recognition test of previously learned auditorily presented items from several associatively-related DRM word lists (Schacter et al., 1996; Schacter, Buckner, Koutstaal, Dale, & Rosen, 1997). Participants made

old–new judgments on items presented during the test phase. Results showed that many brain areas, including the medial temporal lobe (MTL)—comprising the hippocampus and neighboring cortex, anterior prefrontal and orbitofrontal areas, insular cortex, lateral parietal cortex, and visual cortex—that were activated during correct recognition judgments ("hits") were also activated during false recognition ("false alarms"). This indicated that false memory judgments relied, in large part, on similar brain mechanisms as those for true memory judgments. Evidence for differential brain activity for true and false recognition in DRM items was inconclusive, with some studies reporting higher activity for hits than for false alarms (Kim & Cabeza, 2007), but other studies showing no differential effect (Cabeza, Rao, Wagner, Mayer, & Schacter, 2001; Paz-Alonso, Ghetti, Donohue, Goodman, & Bunge, 2008; Schacter et al., 1997). Thus, the neural mechanisms underlying false recollections largely overlap with those of true recollections.

In these studies, false memories were inferred from the old–new judgments that participants made. Such judgments provide little information about the subjective mental experience during recollections. To ascertain whether false memory judgments are based on memory processes rather than response bias, participants can be asked about the quality of their recollective experience. People tend to discriminate true from false memories by judging the amount of perceptual or contextual detail or "evidence" that is associated with recollections (M. K. Johnson et al., 1993; Roediger & McDermott, 2000). True recollections tend to have more recollective detail than false recollections. To obtain information about recollective detail, participants must provide a memory judgment based on perceptual or contextual detail in addition to old-true judgments (Tulving, 1985). The *Remember/Know paradigm* (R/K) has been used to address this issue (Rajaram, 1993; Roediger & McDermott, 1995; Tulving, 1985). In this paradigm, participants make Remember/Know judgments for those items they report to have recognized from a previously studied list (i.e., old/new judgment), which provide information about whether participants have explicit knowledge of item features or contextual details ("recollection") or have a more implicit sense of the experience in memory ("familiarity").

Generally, participants report less phenomenological detail with falsely recognized items compared with accurately recognized items. That is, false memory judgments are associated with more Know than Remember responses, which indicates that false memories possess less perceptual detail or associative context (Rajaram, 1993; Tulving, 1985). However, participants are more likely to make Remember endorsements with false memories if their recollections come to mind more easily or are made more distinctive from other memories (Roediger & McDermott, 1995). One way that this effect is achieved could be through increased sensory or mnemonic processing prior to recognition testing.

Functional neuroimaging studies that used the R/K paradigm have shown greater MTL activity during memory retrieval for Remember judgments compared with Know judgments (Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000; Wheeler & Buckner, 2004). This finding fits well with the long recognized role of the MTL in episodic memory formation (Milner, Squire, & Kandel,

1998; Scoville & Milner, 1957; Squire, 1992). Several decades of clinical research has shown dense anterograde and retrograde amnesia for episodic events in patients with bilateral hippocampal lesions (Milner et al., 1998; Scoville & Milner, 1957). At the same time, research in rodents has shown that hippocampal neural cells also encode for contextual features in perception and memory, such as space (Burgess, Maguire, & O'Keefe, 2002) and time (Eichenbaum, 2014), which indicates that the MTL is particularly associated with encoding associative information. The MTL's place within the brain's anatomical architecture further supports the associative processing role of the MTL: It is strongly interconnected with many cortical and subcortical structures (Burgess et al., 2002; Squire, 1992), which suggests that it may be well suited to bind inputs from various neural sources that represent different mental experiences. This description of the MTL as an associative module fits well with the notion that episodic memories are inherently associative, that is, episodic memories are context-dependent mental experiences (Moscovitch et al., 2005). Indeed, several fMRI studies showed greater MTL activity when participants correctly recalled items in context, such as the source of an item or the spatial or temporal configuration of a set of items, compared with recalling or recognizing single items (Cansino, Maquet, Dolan, & Rugg, 2002; Weis et al., 2004). The MTL may thus contribute to the activation of contextual associations that underlie Remember endorsements of true recollections.

In addition, Remember responses for correctly recognized items are also related to increased activity in the sensory perception cortex in comparison with Know responses for such items (Wheeler & Buckner, 2004). This suggests that the act of remembering recruits perceptual content stored in memory, with the degree of sensory cortical activity during memory retrieval possibly indicating the perceptual vividness of the recollection. More generally, this finding points to the suggestion that memory for perceptual details is stored in those brain areas that encode perceptual information (Buckner & Wheeler, 2001). Indeed, human neuroimaging studies have shown that recalling visual objects from memory activates early and higher order visual cortex, including object-related brain areas (Slotnick & Schacter, 2006; Wheeler, Petersen, & Buckner, 2000), whereas recalling sounds or tunes activates content-related auditory cortex (Halpern & Zatorre, 1999; Linden et al., 2011). These findings appear in line with a series of case reports (Penfield & Perot, 1963) in which patients who were to undergo brain surgery to treat intractable epilepsy received intracortical brain stimulation to sensory cortical sites while they were awake and conscious. Local stimulation of the visual cortex resulted in several patients reporting seeing faces or scenes from previous experiences, comparable to a hallucinatory-like experience. Likewise, stimulation of the auditory cortex resulted in some patients reporting hearing voices, sounds, or music from previous experiences. In recent years, noninvasive brain stimulation techniques, such as transcranial magnetic stimulation, have been applied to visual cortical sites in healthy participants while they performed a memory task (Silvanto, Muggleton, & Walsh, 2008; van de Ven & Sack, 2013). When administered over the occipital cortex during moments of short-term memory retention, participants showed impaired subsequent recognition performance, indicating

that brain stimulation interfered with memory representations in visual cortex (Cattaneo, Vecchi, Pascual-Leone, & Silvanto, 2009; van de Ven, Jacobs, & Sack, 2012).

These latter findings are controversial because they contrast the more traditional, modular view of brain processing in which sensory brain areas are passive encoders of incoming sensory information and memories are stored in MTL and higher order processing areas. Instead, there is strong evidence that the neural representation of memory is distributed across the entire brain, with functionally specialized brain areas supporting encoding as well as storage of respective content or cognitive information. Activation of parts of the distributed neural traces could spread to other parts of the memory representation, that is, reactivate the neural network that represents the memory of the experience (Kandel, Dudai, & Mayford, 2014). These findings provide further support for postulations about how memories are formed that were made more than a century ago (James, 1890).

However, other studies have shown that the link between sensory cortical activity and memory retrieval may not be straightforward. In one study (Slotnick & Schacter, 2004), participants learned a list of visual abstract items, which activated the early visual cortex. During the testing phase, visual cortical activity for correctly recognized items was not different than for missed items. Another study used a method to analyze distributed patterns of activity (for more information on this method, see later) that were measured when participants provided R/K-like judgments (J. D. Johnson, McDuff, Rugg, & Norman, 2009). Results showed that the pattern of brain activity during retrieval was similar to that of encoding for both types of judgment, indicating that reactivation occurred irrespective of participants' memory decisions. These findings suggest that putative memory traces in the sensory cortex may not always be used in—or be accessible for—memory judgments.

Neural Correlates of False Memory Creation: A Role for Encoding

Many of the aforementioned studies focused on false memory retrieval. In fact, false memory judgments may also stem from processing mechanisms prior to retrieval, that is, during the encoding and storage of perceptual experiences in memory. As already stated by William James, the functional role of encoding is to facilitate the "liability to recall" (James, 1890). In other words, encoding mechanisms can facilitate the endurance of a memory. Importantly, the memory encoding mechanisms are constructive in nature and thus play an active role in how information is processed. Manipulation of encoding processes can thus alter the fate of memories. Faulty encoding of current mental experiences and their associations could result in memories that are unrelated to the initial experience. Indeed, there are ample examples of behavioral studies that demonstrated an increased likelihood for false memory judgments under certain encoding conditions (Gallo & Roediger, 2002; Howe et al., 2009; Roediger & McDermott, 2000).

It is well known that the hippocampus and other MTL structures are important for memory encoding. Hippocampal damage can lead to complete loss of

the ability to form new memories (Milner et al., 1998). Further, functional imaging studies in healthy participants have shown that MTL activity increases under contextual encoding conditions. For example, tasks that require participants to pair or bind items together during encoding activate the hippocampus more strongly compared with when participants encode individual items (Davachi & Wagner, 2002; Giovanello, Schnyer, & Verfaellie, 2004). The MTL has also been shown to process temporal contexts, such as temporal proximity or temporal regularity between visual items, during encoding (Ezzyat & Davachi, 2014; Staresina & Davachi, 2009; Tubridy & Davachi, 2011). Furthermore, the strength of MTL activity during encoding predicts subsequent retrieval success (Davachi & Wagner, 2002; Kim & Cabeza, 2007), which indicates that the associative role of the MTL is important in facilitating the endurance of memories. Further, it is well known that attention facilitates encoding of sensory events (Desimone & Duncan, 1995; Kastner & Ungerleider, 2000; Treisman & Gelade, 1980). FMRI studies have shown increased sensory cortical activity when selectively attending visual or auditory objects, resulting in more information processing to optimize task performance and decision making (see Kastner & Ungerleider, 2000).

Not all items presented during encoding will be processed in the same way. Changes in mental or neurophysiological states at different moments in time will alter how individual items are processed. To better understand the brain processes that underlie the subsequent fate of memories, investigators have used trial sorting approaches to group encoding trials according to retrieval success. In this approach, which has also been termed the *subsequent memory paradigm* (Paller & Wagner, 2002), encoding trials are sorted posthoc according to participants' subsequent retrieval success, thereby contrasting brain activity of encoding trials that resulted in successful retrieval to those encoding trials that resulted in failed retrieval (see Figure 5.3 for a schematic overview). Brain areas that show differential encoding activity for items that are later successfully retrieved could then be important in facilitating memory processes. Furthermore, the approach allows investigation of how moment-to-moment variations in mental or physiological conditions during encoding increase the endurance of memories within participants, such as changing levels of attention, working memory performance, or behavioral goals (for a review, see Paller & Wagner, 2002).

The relevance of subsequent memory sorting of encoding trials has been used in a number of fMRI studies that used a *reality monitoring paradigm*, in which during the encoding phase, participants either saw pictures of objects or had to mentally imagine the object that was cued only by name. Encoding was implicit, that is, participants made semantic decisions about the seen or imagined object without instructions to remember the object. In a subsequent recognition test, participants had to make an old/new judgment and judge the source of the item, that is, if they had seen it ("external source") or had imagined it ("internal source"). Correct source judgments constituted an external source judgment for items they had previously seen. A false memory constituted an external source judgment for items they had actually previously imagined. FMRI studies that used this paradigm showed stronger activity in sensory areas for items that were seen during

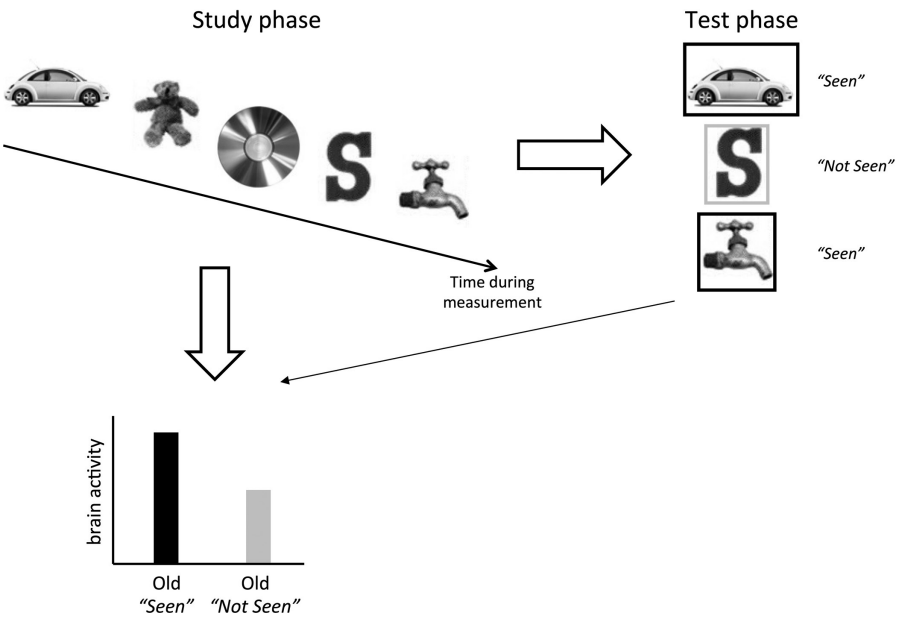

Figure 5.3 Trial sorting approach to analyze how encoding-related activity predicts subsequent successful memory retrieval. See main text for more details.

the encoding phase compared with imagined items (Kensinger & Schacter, 2005), a finding that is in line with other studies on mental imagery of visual objects (Goebel, Khorram-Sefat, Muckli, Hacker, & Singer, 1998; Kosslyn, Thompson, Kim, & Alpert, 1995; O'Craven & Kanwisher, 1999). However, imagined items that subsequently led to a false memory judgment showed more activity in perceptual processing areas during encoding than imagined items that did not lead to false memories (Gonsalves & Paller, 2000; Gonsalves et al., 2004; Kensinger & Schacter, 2005). A comparable finding was also reported for an auditory-based reality monitoring task, in which increased left inferior frontal cortex activity during auditory imagery of words was associated with subsequent judgments that these words were acoustically presented to participants during encoding (Sugimori, Mitchell, Raye, Greene, & Johnson, 2014). These findings suggest that increased sensory cortical activity during mental imagery of objects during encoding could increase the likelihood of subsequent source misattributions. Possibly, increased sensory cortex during encoding could indicate enhanced processing of perceptual features, which are subsequently encoded into memory. In other words, the subjective vividness in a recollected experience may contribute to the decision about whether that recollection is predicated on a sensory-based experience. There is some empirical support for this postulation, with one fMRI study showing increased sensory cortical activity for recollections that were erroneously judged to be based on sensory experiences when they were in fact mentally

imagined during encoding (Kensinger & Schacter, 2006). Thus, it is possible that a false memory constitutes too rich or vivid a recollection of perceptual features.

In sum, fMRI studies of false memories have shown a substantial overlap between the neural networks that support true and false memories. This is in accordance with the notion that true and false memories arise from the same cognitive operations. Further, fMRI studies of memory encoding showed that overly increased activity in brain areas associated with perception and higher cognitive functions increase the likelihood of subsequent false memory judgments. Thus, false memories could arise when the encoding of the respective mental events incidentally becomes more similar to that of encoding of true memories, which suggests that memory encoding processes—including sensory perception, attention, associative encoding, and other cognitive functions—may play a crucial role in false memory generation.

Attentional Selection and Control of Memory Formation and Retrieval

In addition to the MTL and sensory cortical areas, activity in higher-order cortical areas in the frontal and parietal lobes has also been associated with true and false memories. Studies in patients have shown that damage to the prefrontal cortex (PFC) may lead to various memory problems, including deficits in working memory (Owen, Downes, Sahakian, Polkey, & Robbins, 1990; Ranganath & Knight, 2002), memory recall (Gershberg & Shimamura, 1995; Shimamura, Janowsky, & Squire, 1990), and source misattributions or confusions—arguably types of false memory (Janowsky, Shimamura, & Squire, 1989; Moscovitch & Melo, 1997). FMRI of healthy participants has shown increased PFC activity during retention of items in working memory (Courtney, Ungerleider, Keil, & Haxby, 1997; Jiang, Haxby, Martin, Ungerleider, & Parasuraman, 2000; Linden et al., 2003; Todd & Marois, 2004), as well as directing attention to mental representations in short- and long-term memory (Blumenfeld & Ranganath, 2007; Ranganath, Johnson, & D'Esposito, 2003). Generally, the PFC has been associated with many higher order cognitive functions, including executive control (Courtney et al., 1997; Munk et al., 2002), error monitoring (Van Veen & Carter, 2002), and planning and attentional selection (Nelissen, Stokes, Nobre, & Rushworth, 2013; Zanto & Gazzaley, 2009; Zanto, Rubens, Thangavel, & Gazzaley, 2011), which indicates that the PFC exerts cognitive control over memory functions. Further, PFC areas are anatomically and functionally interconnected to hippocampal structures, thereby exerting control over the associative process of items in memory.

The PFC has a heterogenous organization that includes several functional and anatomical subdivisions, many of which remain to be conclusively mapped to cognitive functions. There is some evidence that different PFC areas contribute differentially to memory of item or source information. The dorsolateral PFC (DLPFC) shows greater activity for encoding the relational features between items than for encoding item identity (Blumenfeld, Parks, Yonelinas, & Ranganath, 2011; Munk et al., 2002; Murray & Ranganath, 2007). This could be related to

the putative functional role of the DLPFC in the executive functions of working memory, particularly the manipulation and prioritization of information in memory. The ventrolateral PFC (VLPFC) has been found to be related to the selection of goal-relevant information or the inhibition of goal-irrelevant information (Blumenfeld & Ranganath, 2007). Further, VLPFC activity during encoding may predict subsequent memory recognition for items that were bound during encoding (Staresina & Davachi, 2006; Staresina, Gray, & Davachi, 2009; Sugimori et al., 2014). During memory testing, there is greater activity in both the DLPFC and VLPFC for contextually processed items than for item identity, but the contributions to memory retrieval may be different. There is some evidence for a hemispheric lateralization to contextual memory processing in the PFC, with the left PFC related to retrieval of contextual (source) memory and the right PFC related to retrieval of item memory (Ranganath, Johnson, & D'Esposito, 2000; Rugg, Fletcher, Chua, & Dolan, 1999; Slotnick, Moo, Segal, & Hart, 2003). An fMRI study using the DRM paradigm found increased activity of the left VLPFC in both true and false recognition (Paz-Alonso et al., 2008), which fits with the suggestion that the left PFC supports contextual memory processing. However, it is also possible that lateralized differentiation of the PFC may be associated with how information is monitored in memory, with the left PFC being associated with strategic monitoring of specific information and the right PFC being associated with more heuristic memory judgments (Blumenfeld & Ranganath, 2007). This differentiation has some overlap with the recollection versus familiarity ratings in R/K paradigms, for which there is also some evidence for a respective left versus right PFC differentiation (Ranganath et al., 2000).

Parietal areas have been associated with a variety of perceptual and higher order cognitive functions, including top-down attentional control (Hopfinger et al., 2000), mapping salience of perceptual items (Gottlieb, Kusunoki, & Goldberg, 1998), comparison of an internal template with perceptual evidence (Ploran et al., 2007), and decision making (Platt & Glimcher, 1999). Many of these functions also interplay with memory functions (Ranganath & Rainer, 2003; Wagner, Shannon, Kahn, & Buckner, 2005), although the exact mechanism of interaction remains unclear.

Several studies showed increased activity in lateral parietal areas, particularly at and around the intraparietal sulcus (IPS) during memory processes, including number of items retained in (short-term) memory (Todd & Marois, 2004), successful encoding of stimulus features into memory (Uncapher & Wagner, 2009), and retrieval of episodic events from memory (Buckner & Wheeler, 2001). However, contrary to sensory cortical areas, the parietal cortex does not seem to be specifically involved in memory for sources or other contexts per se. Rather, the IPS may serve attentional allocation functions that prioritize internal representations or perceptual features for further processing (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008), much like attentional enhancement of incoming sensory stimuli from an external source. This internal attentional role of the parietal cortex could also be associated with the integration of perceptual and contextual information in order to guide decision-making (Platt & Glimcher, 1999; Ploran

et al., 2007). Neuropsychological studies in patients with parietal lobe damage provide support for this account, where these patients show no impairment in source memory judgments, but do show lower confidence in their judgments (Simons, Peers, Mazuz, Berryhill, & Olson, 2010).

At the same time, medial parietal areas, including posterior cingulate cortex and the precuneus, have been associated with processing of information relevant to the "self" (Cavanna & Trimble, 2006; Rameson, Satpute, & Lieberman, 2010; Spreng, Mar, & Kim, 2009). Medial parietal areas are activated during encoding and retrieval of autobiographical information (Buckner & Wheeler, 2001; Spreng et al., 2009). The integrative account of lateral parietal cortex could be extended to include the integration of perceptual information from the external world with self-referential information in order to optimize decision making in favor of one's own goals and beliefs. Finally, the parietal cortex is strongly interconnected with MTL structures, including the hippocampus (Sestieri, Corbetta, Romani, & Shulman, 2011; Vincent et al., 2006), which further supports the integrative account of the parietal cortex in memory formation and retrieval.

In summary, lateral and medial frontal and parietal areas underlie cognitive control and attentional selection that support memory formation and retrieval. In this manner, these areas control and guide content-related and associative processing in perceptual and MTL areas, respectively. Attentional processes prioritize target and contextual information during an experience, related to information from the external world as well as self-referential, which are then bound in memory through activity in the hippocampus and perceptual content areas. However, these processes would not act differently for true and false memories, thereby making it difficult for fMRI to detect false memory-related activity in these areas.

Recent Advances: Mapping Distributed Memory Representations

The studies described have examined how strength or amplitude of brain activity differs for various memory encoding and retrieval conditions. Recently, computational methods for fMRI analysis have been developed that go beyond the mapping of amplitudes and analyze how distributed patterns of brain activity are associated with classes of mental states (Norman, Polyn, Detre, & Haxby, 2006). With this approach, a computer algorithm first learns how best to classify two or more brain states according to the patterns of activity that are associated with these states. Then, the classification information is used to predict the classification of the same brain states with new data. The approach has been referred to by various names, such as multivoxel pattern analysis (MVPA) or, more colloquially, "brain reading," as the method has been able to differentiate brain activity between perceptual or mental conditions that are otherwise not detectable with standard methods (Haxby, Connolly, & Guntupalli, 2014). For example, several research groups were able to dissociate brain activity in the sensory cortex between different subliminal perceptions, which could not be performed with standard approaches (Haynes & Rees, 2005; Kamitani & Tong, 2005). More sensational demonstrations of this method have been the classification of brain

activity patterns of particular sensory content in dream states (Horikawa, Tamaki, Miyawaki, & Kamitani, 2013) and optimizing how a participant learns a visual skill without actually seeing the stimuli that must be learned (Shibata, Watanabe, Sasaki, & Kawato, 2011). An in-depth treatment of this application is beyond the scope of this chapter, and the reader is referred to several excellent reviews on the topic (Haxby et al., 2014; Norman et al., 2006; Rissman & Wagner, 2012).

MVPA's sensitivity to distributed patterns of activity makes it very interesting for use in classifying brain states related to memory functions or phenomenal content (Rissman & Wagner, 2012). For example, MVPA has been used to reveal the contents of short-term memory in visual cortex when participants retained previously seen visual information in mind (Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009). Moreover, MVPA of brain activity during short-term memory retention shows that parietal cortex does not represent perceptual contents, but rather represents information that is relevant for stimulus matching and decision making (Christophel, Hebart, & Haynes, 2012).

In recent years, Wagner and colleagues used MVPA to address the issue of whether false memories can be reliably detected in the absence of participant reports about their memories (Rissman & Wagner, 2012). In one study (Kuhl, Rissman, Chun, & Wagner, 2011), the authors investigated brain activity when two memories competed for retrieval. Participants first learned paired associations between words and pictures of either faces or scenes. Symbolically, learning followed *AB* pairing, comprising memory cues *A* and paired associates *B*. During the experiment, however, the association was altered, such that participants learned to associate the same words with new pictures (*C*), following an *AC* pairing. Specifically, if in the *AB* pairing the associate was a face, then in the *AC* pairing the associate was a scene. Thus, during the experiment the same cue became associated with two different associates in memory. This led to competition of retrieval between the associates when the cue was presented. In addition, participants learned noncompeting cue–associate pairings (*DE* pairings). During test phases, participants were shown the cue and had to indicate if they had specific (that is, similar to an R/K Remember response) or general knowledge (similar to a Know response) about the associate or if they did not know about the associate.

The choice for faces or scenes as associates was crucial, as it utilized the common finding that anatomically different brain areas process faces (Kanwisher, McDermott, & Chun, 1997) and visual scenes (Epstein, Harris, Stanley, & Kanwisher, 1999). Thus, whether during testing a cue elicits recall of a face or a scene can be inferred from the peak activity in either the brain areas that processes faces or that processes scenes. Further, the authors reasoned that confusion in recalling the associate could result in activity in both face and scene processing areas.

MVPA was used to classify activity for the retrieval of faces or scenes in the MTL and higher order visual processing areas. Results showed MVPA could classify faces versus scenes more accurately for *AB* and *DE* cues during retrieval than for the *AC* cues. The authors interpreted this finding as indicating that recall after *AC* pairing suffered from competition from the previously learned *AB* pairings.

Furthermore, classification accuracy was higher for trials in which participants reported having specific knowledge about the associate (similar to Remember), compared with reports of having general information (similar to Know), which fits with differential brain activity for recollection versus familiarity ratings. Thus, these findings show how patterns of brain activity change as a function of competition between items in memory. Further, they suggest that it could be possible to map false memories—as misassociations between different mental events—by the distributed pattern of activity, given that one knows the original source material on which memories were based.

However, it was subsequently shown that using MVPA to predict memory-related brain states may be limited to particular situations. In another study (Rissman, Greely, & Wagner, 2010), the authors used MVPA to classify previously seen from novel faces during an *explicit* recognition task, in which participants gave an old/new judgment to each presented face. Classification accuracy was above 80% for hits versus correct rejections, indicating good separability of whether participants recognized a face. Interestingly, MVPA classification seemed to rely mostly on frontal and parietal areas, with comparatively little contribution from MTL regions. More importantly, the authors conducted a second experiment in which participants *incidentally* encoded faces and then completed an implicit recognition task. This scenario is of particular interest, as it mimics real-life cases in which participants are not aware that their memory will be tested in the future. In this second experiment, MVPA classification dropped to chance performance, indicating that there was no reliable separation between patterns of brain activity for previously seen versus novel faces in implicit memory scenarios. Thus, although it is possible to harvest more information from the brain when participants make explicit memory judgments, fMRI is insensitive to whether a mental experience is implicitly associated with a perceptual memory.

Recent Advances: Manupulating Memories

Ultimately, the brain mechanisms for true and false memories are arguably best described at the level of neural interactions. Although recent developments in fMRI technology have made it possible to measure brain activity at very high spatial and temporal resolution, the inherent workings of fMRI preclude direct measurement of neural activity. For this endeavor, neurophysiological and molecular animal research currently remains the only option.

Recent advances in animal molecular neuroscience studies have provided an innovative insight into the creation of false memories. More specifically, using the methodology of optogenetics (Fenno, Yizhar, & Deisseroth, 2011), researchers obtain experimental control over the activity of *mnemonically specific* neural populations in the hippocampus and create a false memory of fear in living and freely exploring mice. In optogenetics, researchers shine light through an implanted fiber opic cable on neurons that have been genetically manipulated to express light-sensitive receptors on their membranes. The genetic manipulation

is required because neurons are not inherently programmed to express light-sensitive receptors (the manipulation is done in vivo and without otherwise altering neural functioning). Activation of the light-sensitive receptors by turning on the light source causes neurons to increase or decrease their activity, according to experimental parameters such as the type of receptors and frequency of the light.

Pioneering optogenetics work in mice showed that it is possible to activate hippocampal neurons that had become part of a fear memory. In this work, mice were allowed to explore two environments, where in one of the environments the animal received a shock that caused a fear response and activated hippocampal cells, resulting in the formation of a fear memory (Liu et al., 2012). Hippocampal cells were genetically prepared to express light-sensitive receptors when they showed increased activity. As a result of these manipulations, hippocampal cells that were activated during fear learning expressed light-sensitive receptors. In other words, these hippocampal cells became tagged as a result of memory formation. The animal was then returned to the neutral environment, in which it never received a shock. In this neutral environment, turning on the light source resulted in increased activity of the receptor-expressing hippocampal cells, which resulted in fear responses in the mouse. The animal thus behaved as if it responded to being in the fearful environment. Thus, by controlling the activity of the neural ensemble that supported a fear memory, the mouse could be made to retrieve the fearful memory under experimental control.

This approach was also used to create a false memory in mice (Ramirez et al., 2013). Here, hippocampal cells were functionally tagged for optogenetic manipulation using a neutral environment. Subsequently, the mice were put in a new, fearful environment. Crucially, during fear learning, the previously tagged hippocampal cells that coded for the neutral environment were optogenetically activated, which resulted in activation of the memory of the neutral environment concomittant to experiencing the fearful environment. The coactivation of the two environments resulted in them being associated in memory. When put back in the neutral environment, mice that had undergone the optogenetic treatment showed more fear responses than mice that went through fear conditioning without concomittant optogenetic manipulation. This finding demonstrated that false memories can result from associations between cues of different mental experiences. In other words, these findings provide evidence at the level of neural interactions in support of an associative account of false memory creation (Mitchell & Johnson, 2000; Roediger & McDermott, 2000). How to translate this paradigm to human research is an immediate challenge for future studies.

Neuroscience in the Courtroom

This chapter started by asking whether neuroscience could help those in the judiciary (police, prosectors, jurors, and judges) to distinguish between true and false memories. Although the research we have reviewed has considerable promise, there are a number of limitations to the use of neuroscience in the courtroom at this juncture. It is important to stress that fMRI images have played a role in

some cases regarding the reliability of statements. Specifically, legal professionals have become interested in whether fMRI might be useful in the detection of deception;[1] something which is related to true and false memories. Although we will not elaborate on such cases, we do want to make the point that several words of caution are warranted when neuroscience is used to differentiate between true and false memories.

To begin, an often overlooked fact, by scientists and laypeople alike, is that fMRI results represent the output of carefully designed and controlled experiments in combination with a series of decisions about data acquisition, preprocessing, and analysis (Amaro & Barker, 2006; Logothetis, 2008). As previously described, typical fMRI studies of true and false memories require participants to attend to carefully selected stimulus materials under controlled contextual and environmental parameters. Brain activity is then analyzed as a function of the experimental parameters. In other words, experimenters know, or at least have a fairly good sense, about what the participant saw and what he or she ideally did with the information mentally. Typically, in a courtroom one does not have this level of control or knowledge about what the eyewitness experienced.

A further complication is that in many (but not all) fMRI studies, the presented results are statistical summaries of population effects; that is, they rely on the distribution of brain responses from the sampled participants. This approach gives a description about the general pattern of brain activity but does not (necessarily) provide a description of brain activity in a single participant. Recent methodological advances indicate that it is possible, under certain circumstances, to obtain impressions from fMRI images that indicate what a single participant experienced while the experimenter is blind to the content of the experience. However, while intriguing and scientifically innovative, these findings require participant cooperation and carefully controlled experimental conditions. Furthermore, even under these controlled circumstances, the results are not reliable enough to be held as judicial evidence in court.

Finally, fMRI is inherently correlational, which means that it provides no basis for a broadly causal interpretation of the findings (Logothetis, 2008; Poldrack, 2008). An often made error is to infer a brain or mental state from observing a change in brain activity: *Because there is activity in the hippocampus, I know that the participant is now remembering what previously happened*. This inferential problem has been termed the reverse inference problem, because it reverses the chain of events in the experimental design (Poldrack, 2008). Brain activity in a particular area can show a change in level of activity when a participant reports recognizing a stimulus from a previously shown list of items: *On average there is activity in the hippocampus when I ask the participant to now try and remember what previously happened*. This does not mean that seeing the same change of brain activity in the same brain area at a different moment in time means that the participant is now recalling the same stimulus from memory.

Thus, while fMRI research can provide important insights into how cognitive functions are generally mapped onto the brain, the method is not suitable to provide case-specific judgments about whether a person provides a true or false

memory statement at high accuracy (Schacter & Loftus, 2013). As with many other previously proposed techniques, fMRI does not provide an unbiased and objective window into the mental landscape of an individual person in order to unveil private or subjective mental states.

CONCLUSION

In sum, this chapter has described relevant human neuroimaging research that has provided insights into how false memories arise in the brain. The neural and cognitive mechanisms of false memories are largely the same as those for true memory formation and retrieval. The brain forms memories of perceptual or emotional events through the interaction of widely-distributed networks of functionally specialized brain areas. Sensory processing areas encode and store perceptual content of experiences. Frontal and parietal areas control and monitor stimulus processes by implementing goal-directed selection and processing strategies, and prioritize processing of items and relational features for encoding into as well as retrieval from memory. Medial parietal areas and MTL structures encode the relational features of stimuli or mental experiences into memory. Memory retrieval relies for a large part on reactivation of the encoding pathways, under prefrontal executive control. The hippocampus and other MTL areas can erroneously create associations between unrelated items and contexts during encoding as well as during retrieval. Further, overactivity in the sensory cortex during encoding mentally imagined visual objects can lead to subsequent false memory judgments that these objects were actually seen. It remains to be investigated when and how sensory overactivity comes about during mental imagery, and how it affects relational memory processing (e.g., in the hippocampus) and memory decision making (e.g., in parietal cortex).

NOTE

1. See for example http://www.wired.com/2009/03/noliemri/ or http://www.wired.com/2010/05/fMRI-in-court-update/

REFERENCES

Amaro, E., & Barker, G. J. (2006). Study design in fMRI: Basic principles. *Brain and Cognition*, *60*(3), 220–232.

Blumenfeld, R. S., Parks, C. M., Yonelinas, A. P., & Ranganath, C. (2011). Putting the pieces together: The role of dorsolateral prefrontal cortex in relational memory encoding. *Journal of Cognitive Neuroscience*, *23*(1), 257–265.

Blumenfeld, R. S., & Ranganath, C. (2007). Prefrontal cortex and long-term memory encoding: An integrative review of findings from neuropsychology and neuroimaging. *The Neuroscientist*, *13*(3), 280–291.

Buckner, R. L., & Wheeler, M. E. (2001). The cognitive neuroscience of remembering. *Nature Reviews Neuroscience*, *2*(9), 624–634.

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, *35*(4), 625–641.

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*, *9*(8), 613–625.

Cabeza, R., Rao, S. M., Wagner, A. D., Mayer, A. R., & Schacter, D. L. (2001). Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(8), 4805–4810.

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex*, *12*, 1048–1056.

Cattaneo, Z., Vecchi, T., Pascual-Leone, A., & Silvanto, J. (2009). Contrasting early visual cortical activation states causally involved in visual imagery and short-term memory. *European Journal of Neuroscience*, *30*(7), 1393–1400.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564–583.

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience*, *32*(38), 12983–12989.

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

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, *386*, 608–611.

Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *Journal of Neurophysiology*, *88*(2), 982–990.

Deese, J. (1959). On the prediction of occurence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Eichenbaum, H. (2014). Time cells in the hippocampus: A new dimension for mapping memories. *Nature Reviews Neuroscience*, *15*(October), 732–744.

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, *3*(11), 1149–1152.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, *23*(1), 115–125.

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, *81*(5), 1179–1189.

Fenno, L. E., Yizhar, O., & Deisseroth, K. (2011). The development and applications of optogenetics. *Annual Review of Neuroscience*, *34*, 389–412.

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language*, *47*(3), 469–497.

Gershberg, F. B., & Shimamura, A. P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychologia*, *13*(10), 1305–1333.

Giovanello, K. S., Schnyer, D. M., & Verfaellie, M. (2004). A critical role of the anterior hippocampus in relational memory: Evidence from an fMRI study comparing associative and item recognition. *Hippocampus*, *14*(1), 5–8.

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., & Singer, W. (1998). The constructive nature of vision: Direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal of Neuroscience*, *10*(5), 1563–1573.

Gonsalves, B., & Paller, K. A. (2000). Neural events that underlie remembering something that never happened. *Nature Neuroscience*, *3*(12), 1316–1321.

Gonsalves, B., Reber, P. J., Gitelman, D. R., Parrish, T. B., Mesulam, M.-M., & Paller, K. A. (2004). Neural evidence that vivid imagining can lead to false remembering. *Psychological Science*, *15*(10), 655–660.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*(6666), 481–484.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*, 697–704.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.

Haxby, J. V, Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*, 435–456.

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686–691.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, *3*(3), 284–291.

Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, *340*(6132), 639–642.

Howe, M. L. (2013). Memory development: Implications for adults recalling childhood experiences in the courtroom. *Nature Reviews Neuroscience*, *14*(12), 869–876.

Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language*, *60*(2), 229–251.

AQ: Pls. clarify this; also provide city and publisher.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging. Book* (Vol. 23).

James, W. (1890/1950). *The principles of psychology* (Vol. 1). New York: Dover Publishers, Inc.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia*, *27*(8), 1043–1056.

Jiang, Y., Haxby, J. V, Martin, A., Ungerleider, L. G., & Parasuraman, R. (2000). Complementary neural mechanisms for tracking items in human working memory. *Science*, *287*(5453), 643–646.

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*, *63*(5), 697–708.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28.

AQ: City of publication?

Johnson, M. K., Raye, C. L., Mitchell, K. J., & Ankudowich, E. (2012). The cognitive neuroscience of true and false memories. In R. F. Belli (Ed.), *True and false recovered memories: Toward a reconciliation of the debate* (Vol. 58, pp. 15–52). : Springer.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685.

Kandel, E. R., Dudai, Y., & Mayford, M. R. (2014). The molecular and systems biology of memory. *Cell*, *157*(1), 163–186.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341.

Kensinger, E. A., & Schacter, D. L. (2005). Emotional content and reality-monitoring ability: fMRI evidence for the influences of encoding processes. *Neuropsychologia*, *43*(10), 1429–1443.

Kensinger, E. A., & Schacter, D. L. (2006). Neural processes underlying memory attribution on a reality-monitoring task. *Cerebral Cortex*, *16*(8), 1126–1133.

Kim, H., & Cabeza, R. (2007). Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cerebral Cortex*, *17*(9), 2143–2150.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, *378*, 496–498.

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5903–5908.

Linden, D. E. J., Bittner, R. A., Muckli, L., Waltz, J., Kriegeskorte, N., Goebel, R., . . . Munk, M. H. J. (2003). Cortical capacity constraints for visual working memory: Dissociation of fMRI load effects in a fronto-parietal network. *NeuroImage*, *20*(3), 1518–1530.

Linden, D. E. J., Thornton, K., Kuswanto, C. N., Johnston, S. J., van de Ven, V., & Jackson, M. C. (2011). The brain's voices: Comparing nonclinical auditory hallucinations and imagery. *Cerebral Cortex*, *21*(2), 330–337.

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., & Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, *484*(7394), 381–385.

Loftus, E. F. (1991). Made in memory: Distortions in recollection after misleading information. *Psychology of Learning and Motivation*, *27*(C), 187–215.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist*, *48*(5), 518–537.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878.

Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron*, *20*, 445–468.

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. Craik (Eds.), *The Oxford handbook of memory* (pp. 179–195). New York: Oxford University Press.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638–677.

Moscovitch, M., & Melo, B. (1997). Strategic retrieval and the frontal lobe: Evidence from confabulation and amnesia. *Neuropsychologia*, *35*(7), 1017–1034.

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., . . . Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic

and spatial memory: A unified account based on multiple trace theory. *Journal of Anatomy*, *207*(1), 35–66.

Munk, M. H. J., Linden, D. E. J., Muckli, L., Lanfermann, H., Zanella, F. E., Singer, W., & Goebel, R. (2002). Distributed cortical systems in visual short-term memory revealed by event-related functional magnetic resonance imaging. *Cerebral Cortex*, *12*(8), 866–876.

Murray, L. J., & Ranganath, C. (2007). The dorsolateral prefrontal cortex contributes to successful relational memory encoding. *Journal of Neuroscience*, *27*(20), 5515–5522.

Nader, K., Schafe, G. E., & Ledoux, J. E. (2000). The labile nature of consolidation theory. *Nature Reviews Neuroscience*, *1*(3), 216–219.

Nelissen, N., Stokes, M. G., Nobre, A. C., & Rushworth, M. F. S. (2013). Frontal and parietal cortical interactions with distributed visual representations during selective attention and action selection. *Journal of Neuroscience*, *33*(42), 16443–16458.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430.

O'Craven, K. M., & Kanwisher, N. (1999). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, *28*(10), 1021–1034.

Paller, K. A., & Wagner, A. D. (2002). Observing the transformatioin of experience into memory. *Trends in Cognitive Sciences*, *6*(2), 93–102.

Paz-Alonso, P. M., Ghetti, S., Donohue, S. E., Goodman, G. S., & Bunge, S. A. (2008). Neurodevelopmental correlates of true and false recognition. *Cerebral Cortex*, *18*(September), 2208–2216.

Penfield, W., & Perot, P. (1963). The brain's record of auditory and visual experience. *Brain*, *86*(4), 595–696.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*(2), 175–187.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*(6741), 233–238.

Ploran, E. J., Nelson, S. M., Velanova, K., Donaldson, D. I., Petersen, S. E., & Wheeler, M. E. (2007). Evidence accumulation and the moment of recognition: Dissociating perceptual recognition processes using fMRI. *Journal of Neuroscience*, *27*(44), 11912–11924.

Poldrack, R. A. (2008). The role of fMRI in Cognitive Neuroscience: Where do we stand? *Current Opinion in Neurobiology*, *18*(2), 223–227.

Poldrack, R. A., Fletcher, P. C., Henson, R. N. A., Worsley, K. J., Brett, M., & Nichols, T. E. (2008). Guidelines for reporting an fMRI study. *NeuroImage*, *40*(2), 409–414.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*(1), 89–102.

Rameson, L. T., Satpute, A. B., & Lieberman, M. D. (2010). The neural correlates of implicit and explicit self-relevant processing. *NeuroImage*, *50*(2), 701–708.

Ramirez, S., Liu, X., Lin, P.-A., Suh, J., Pignatelli, M., Redondo, R. L., . . . Tonegawa, S. (2013). Creating a false memory in the hippocampus. *Science*, *341*(6144), 387–391.

Ranganath, C., Johnson, M. K., & D'Esposito, M. (2000). Left anterior prefrontal activation increases with demands to recall specific perceptual information. *Journal of Neuroscience*, *20 (22)*(108), 1–5.

Ranganath, C., Johnson, M. K., & D'Esposito, M. (2003). Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia*, *41*, 378–389.

Ranganath, C., & Knight, R. T. (2002). Prefrontal cortex and episodic memory: Integrating findings from neuropsychology and functional brain imaging. In A. Parker, T. J. Bussey, & E. L. Wilding (Eds.), *The cognitive neuroscience of memory: Encoding and retrieval* (Vol. 1, pp. 83–99). Hove, UK: Psychology Press.

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, *4*(3), 193–202.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724.

Rissman, J., Greely, H. T., & Wagner, A. D. (2010). Detecting individual memories through the neural decoding of memory states and past experience. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(21), 9849–9854.

Rissman, J., & Wagner, A. D. (2012). Distributed representations in memory: Insights from functional brain imaging. *Annual Review of Psychology*, *63*, 101–128.

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*(4), 803–814.

Roediger, H. L., & McDermott, K. B. (2000). Distortions of memory. In E. Tulving & F. I. Craik (Eds.), *Oxford handbook of memory* (pp. 149–162). Oxford: Oxford University Press.

Rugg, M. D., Fletcher, P. C., Chua, P. M., & Dolan, R. J. (1999). The role of the prefrontal cortex in recognition memory and memory for source: An fMRI study. *NeuroImage*, *10*(5), 520–529.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., & Rosen, B. R. (1997). Late onset of anterior prefrontal activity during true and false recognition: An event-related fMRI study. *NeuroImage*, *6*(4), 259–269.

Schacter, D. L., & Loftus, E. F. (2013). Memory and law: What can cognitive neuroscience contribute? *Nature Neuroscience*, *16*(2), 119–123.

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

Schacter, D. L., Reiman, E., Curran, T., Yun, L. S., Bandy, D., Mcdermott, K. B., & Roediger III, H. L. (1996). Neuroanatomical correlates of veridical and illusory recognition memory: Evidence from positron emission tomography. *Neuron*, *17*, 267–274.

Schacter, D. L., & Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron*, *44*, 149–160.

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., & Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Consciousness and Cognition*, *17*(2), 457–67.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *The Journal of Neurology, Neurosurgery & Psychiatry*, *20*(11), 11–21.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214.

Sestieri, C., Corbetta, M., Romani, G. L., & Shulman, G. L. (2011). Episodic memory retrieval, parietal cortex, and the default mode network: Functional and topographic analyses. *Journal of Neuroscience*, *31*(12), 4407–4420.

Shibata, K., Watanabe, T., Sasaki, Y., & Kawato, M. (2011). Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science*, *334*(6061), 1413–1415.

Shimamura, A. P., Janowsky, J. S., & Squire, L. R. (1990). Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. *Neuropsychologia*, *28*(8), 803–813.

Silvanto, J., Muggleton, N. G., & Walsh, V. (2008). State-dependency in brain stimulation studies of perception and cognition. *Trends in Cognitive Sciences*, *12*(12), 447–454.

Simons, J. S., Peers, P. V, Mazuz, Y. S., Berryhill, M. E., & Olson, I. R. (2010). Dissociation between memory accuracy and memory confidence following bilateral parietal lesions. *Cerebral Cortex*, *20*(2), 479–485.

Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research*, *17*, 75–82.

Slotnick, S. D., & Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nature Neuroscience*, *7*(6), 664–672.

Slotnick, S. D., & Schacter, D. L. (2006). The nature of memory related activity in early visual areas. *Neuropsychologia*, *44*(14), 2874–2886.

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (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.

Squire, L. R. (1992). Memory and the hippocampus : A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195–231.

Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *Journal of Neuroscience*, *26*(36), 9162–9172.

Staresina, B. P., & Davachi, L. (2009). Mind the gap: Binding experiences across space and time in the human hippocampus. *Neuron*, *63*, 267–276.

Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex*, *19*(May), 1198–1207.

Sugimori, E., Mitchell, K. J., Raye, C. L., Greene, E. J., & Johnson, M. K. (2014). Brain mechanisms underlying reality monitoring for heard and imagined words. *Psychological Science*, *25*(2), 403–413.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751–754.

Tootell, R. B. H., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From retinotopy to recognition. *Trends in Cognitive Sciences*, *2*(5), 174–183.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Tubridy, S., & Davachi, L. (2011). Medial temporal lobe contributions to episodic sequence encoding. *Cerebral Cortex*, *21*(2), 272–280.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, *26*, 1–12.

Tulving, E., & Craik, F. I. (Eds.). (2000). *The Oxford handbook of memory*. Oxford: Oxford University Press.

Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: Insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiology of Learning and Memory*, *91*(2), 139–154.

van de Ven, V., Jacobs, C., & Sack, A. T. (2012). Topographic contribution of early visual cortex to short-term memory consolidation: A transcranial magnetic stimulation study. *Journal of Neuroscience*, *32*(1), 4–11.

van de Ven, V., & Sack, A. T. (2013). Transcranial magnetic stimulation of visual cortex in memory: Cortical state, interference and reactivation of visual content in memory. *Behavioural Brain Research*, *236*, 67–77.

Van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology and Behavior*, *77*(4-5), 477–482.

Vincent, J. L., Snyder, A. Z., Fox, M. D., Shannon, B. J., Andrews, J. R., Raichle, M. E., & Buckner, R. L. (2006). Coherent spontaneous activity identifies a hippocampal-parietal memory network. *Journal of Neurophysiology*, *96*(6), 3517–3531.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445–453.

Weis, S., Specht, K., Klaver, P., Tendolkar, I., Willmes, K., Ruhlmann, J., . . . Fernández, G. (2004). Process dissociation between contextual retrieval and item recognition. *Neuroreport*, *15*(18), 2729–2733.

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience*, *9*(7), 971–978.

Wheeler, M. E., & Buckner, R. L. (2004). Functional-anatomic correlates of remembering and knowing. *NeuroImage*, *21*(4), 1337–1349.

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*(20), 11125–11129.

Zanto, T. P., & Gazzaley, A. (2009). Neural suppression of irrelevant information underlies optimal working memory performance. *Journal of Neuroscience*, *29*(10), 3059–3066.

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience*, *14*(5), 656–661.