Neuroimaging of True, False, and Imaginary Memories

4 authors, including:

Daniel L Schacter
Harvard University
589 PUBLICATIONS   **86,848** CITATIONS

Brendan Gaesser
University at Albany, The State University of New York
19 PUBLICATIONS   **750** CITATIONS

Draft to appear in: L. Nadel and W. Sinnott-Armstrong (Eds.) *Memory and Law: Perspectives from Cognitive Neuroscience*. New York: Oxford University Press.

Neuroimaging of True, False, and Imaginary Memories:

Findings and Implications

Daniel L. Schacter, Jon Chamberlain, Brendan Gaesser, and Kathy D. Gerlach

Department of Psychology, Harvard University, Cambridge, MA 02138

Correspondence:

Daniel L. Schacter

Department of Psychology

Harvard University

Cambridge, Massachusetts 02138

phone: (617) 495-3856

fax: (617) 496-3122

e-mail: dls@wjh.harvard.edu

In his 2004 book *Against All Enemies*, Richard A. Clarke – former counter-terrorism chief in the Clinton and Bush administrations – shared his recollections of how administration officials responded to the events of September 11, 2001 (Clarke, 2004). He recalled, for example, that the Secret Service asked for fighter escorts to protect Air Force One, and that his national security colleague Franklin Miller urged Secretary of Defense Donald Rumsfeld to take a helicopter out of the Pentagon. Other administration officials challenged these recollections. Franklin Miller recollected that he himself had asked Condoleezza Rice about whether to call up fighter escorts, and further recalled that she told him to go ahead with their deployment. Miller also stated that he never spoke with Secretary Rumsfeld on September 11th. Miller further opined that while Clarke's recollections would "make a great movie", they did not reflect the reality of what happened that day (Schacter, 2004).

Conflicting recollections of the same event are not uncommon in the courtroom. Consider, for example, the contrasting memories that emerged in testimony concerning the July 2005 death of Jean Charles de Menezes, an innocent man who was fatally shot by London police in a subway station. De Menezes was misidentified by police as one of several men who had carried out a failed bombing attempt the previous day:

Firearms officers recalled running on to the Underground platform at Stockwell and challenging de Menezes by shouting "Armed Police," before shooting him seven times in the head. But 17 civilian witnesses could not remember such a thing being said. The police said that the electrician had stood up and walked "aggressively" towards them, but some witnesses do not remember him getting up from his seat. Everyone recalled a slightly different sequence of events, even when it came to such basic facts as the number of bullets fired or the clothes de Menezes was wearing (O'Connell, 2008).

Such conflicting recollections could reflect willful distortion on the part of some or all parties, but are more likely attributable to well-known imperfections of memory that render people vulnerable to various kinds of forgetting and distortion (Schacter, 2001, 2004). Memory

errors are also highly relevant to eyewitness testimony: it is well known that eyewitnesses are prone to memory distortion, sometimes reporting highly confident but inaccurate memories that can have a large influence in the courtroom (e.g., Cutler & Penrod, 1995; Loftus, 1979; Semmler & Brewer, 2010; Wells & Olson, 2003). Indeed, faulty eyewitness testimony was a key factor in approximately 75% of the first hundred individuals who were exonerated by DNA evidence after being convicted of crimes they did not commit (Scheck, Neufeld, & Dwyer, 2000; Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998). These observations raise a question of great practical and theoretical interest: Is there any way to determine definitively whose recollection is accurate and whose is wrong?

The question of whether it is possible to distinguish between true and false memories has long been of interest to psychologists; it achieved special prominence and urgency during the 1990s in relation to the heated controversy concerning the accuracy of recovered memories of childhood sexual abuse (cf., Jacobs & Nadel, 1998; Loftus, 1993; Loftus & Davis 2006; McNally & Geraerts, 2009; Pendergrast, 1995; Schacter, 1996). Many cases of recovered memories that ended up in the courtroom involved conflicting recollections between accusing individuals who claimed to have recalled long-repressed memories of childhood abuse and accused individuals who denied that the abuse ever occurred, thereby posing a difficult challenge for the legal system (Loftus & Ketcham, 1994).

Psychologists have focused on attempting to distinguish between true and false memories based on their qualitative characteristics, asking such questions as whether true memories are more vivid or detailed than false memories (for review, see Bernstein & Loftus, 2009). During the past 15 years, however, another approach to the issue has emerged that focuses on measuring brain activity. Researchers taking this approach have relied on functional neuroimaging

techniques that can accurately localize changes in brain activity associated with various aspects of memory, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), in an attempt to pinpoint specific brain regions that may be differentially associated with true versus false memories. In the present chapter, we discuss this research that has attempted to distinguish brain activity associated with true and false memories using PET and fMRI (for review of related research using event-related potentials (ERPs), an electrophysiological technique that measures changes in voltage topography on the scalp over time, see Schacter and Slotnick (2004); here, we will only note briefly ERP studies relevant to our key themes).

Some strong claims have already been made for the viability of using data generated by neuroimaging techniques to distinguish truth from fiction in settings such as the courtroom. For example, the company No Lie MRI offers the following claim on its website (http://www.noliemri.com):

New Truth Verification Technology

No Lie MRI, Inc. provides unbiased methods for the detection of deception and other information stored in the brain.

The technology used by No Lie MRI represents the first and only direct measure of truth verification and lie detection in human history!

This claim appears to be primarily focused on distinguishing truthful testimony from intentional deception, which is a related though distinct problem from distinguishing between true and false memories: lies are generated with an intent to deceive, whereas people believe that they are telling the truth when they report false memories. Nonetheless, the claim is a strong one. In light of such claims, and with increasing interest in the possibility of applying neuroimaging techniques to courtroom settings, we believe that it is important to critically assess the state of

our knowledge concerning both brain-based lie detection (for reviews, see Abe, 2009; Greely & Illes, 2007; Spence & Kaylor-Hughes, 2008) and brain-based attempts to distinguish true and false memories. We focus on the latter task here.

The chapter consists of three main sections. The first section will discuss attempts to use PET and fMRI to distinguish between true and false memories under controlled laboratory conditions. We will discuss primarily work conducted in our own laboratory, but we will also attempt to relate our findings to those from other labs. Second, we will discuss a recent and related line of work that compares the neural underpinnings of actual memories of past experiences with imaginary experiences of events that might occur in the future, again focusing on work from our lab but also relating this work to research from other labs. Finally, in the third and concluding section of the chapter we will consider limitations of the research we have discussed along with its possible implications for the courtroom.

Neuroimaging of true and false memories

Neuroimaging studies have attempted to distinguish between true and false memories either by measuring brain activity at the time of retrieval, or by measuring brain activity during encoding and asking whether any aspects of encoding-related brain activity predict whether subsequent memories are accurate or distorted. In legal contexts, the potential use of neuroimaging techniques to distinguish between true and false memories would appear to be restricted primarily if not entirely to the time of retrieval. Therefore, we limit our discussion to studies that have measured brain activity during retrieval (for encoding-based studies, see Aminoff, Schacter, & Bar, 2008; Dennis, Kim, & Cabeza, 2007; Garoff, Slotnick, & Schacter, 2005; Gonsalves & Paller, 2000; Gonsalves, Reber, Gitelman, Parrish, Mesulam, & Paller, 2004; Kensinger & Schacter, 2005a; Kim & Cabeza, 2007a; Okado & Stark, 2005).

Most neuroimaging attempts to distinguish between true and false memories have been conducted in the context of what has been termed the *sensory reactivation hypothesis*: the idea that true memories are accompanied by retrieval of more sensory/perceptual details than false memories, which in turn reflect the reactivation of sensory/perceptual encoding processes that were engaged during the establishment of true but not false memories (for further discussion, see Schacter, Norman, & Koutstaal, 1998; Schacter & Slotnick, 2004). The hypothesis originated from behavioral studies that indeed revealed evidence for greater retrieval of sensory/perceptual details during true rather than false memory retrieval (e.g., Johnson, Foley, Suengas, & Raye 1988; Marche, Brainerd, & Reyna, 2010; Mather, Henkel, & Johnson 1997; Norman & Schacter, 1997; Schooler, Gerhard, & Loftus, 1986). This hypothesis naturally leads to the question of whether neural activity accompanying true recognition, compared with false recognition, shows signs of sensory reactivation.

Evidence for sensory reactivation in PET and fMRI studies   Our laboratory initiated neuroimaging investigations of the relation between true and false memories in the mid-1990s. A necessary condition for conducting such investigations is the availability of an experimental paradigm that can induce robust, subjectively compelling false memories that participants believe to be true memories of past experience. Further, it is crucial in neuroimaging studies to obtain sufficient numbers of observations in order to generate a reliable signal that can be distinguished statistically from noise. Therefore, a suitable experimental paradigm must yield a large enough number of false memories to permit a meaningful statistical analysis.

We therefore turned to a then-recently described paradigm by Roediger and McDermott (1995) that produces extremely high levels of a phenomenon known as *false recognition*, where subjects incorrectly claim that a novel item has been encountered earlier in an experiment. False

recognition is typically inferred when participants make "old" responses to novel items that are conceptually or perceptually related to previously studied items -- that is, when the level of false alarms to related novel items is higher than the "baseline" level of false alarms to unrelated novel items. To produce robust false recognition, Roediger and McDermott modified a procedure developed earlier by Deese (1959) in which subjects hear lists of associated words (e.g., *candy, sour, sugar, bitter, good, taste, tooth*, etc.) that all converge on a nonpresented "theme word" or false target (e.g., *sweet*). Roediger and McDermott reported extremely high levels of false recognition (e.g., 80%) to the theme words across a variety of word associate lists. The level of false recognition responses to the false target was indistinguishable from the hit rate to studied items, and the false recognition responses were accompanied by very high confidence, comparable to confidence associated with true recognition responses. These initial observations were confirmed and extended to numerous subsequent studies using the "DRM" paradigm, which have delineated various cognitive properties of this potent false recognition effect (for a review, see Gallo, 2006).

In our first neuroimaging study of true and false recognition, PET scans were carried out while subjects performed a recognition test following study of various DRM associate lists (Schacter, Reiman, Curran, Sheng Yun, Bandy, McDermott, & Roediger, 1996). After auditory presentation of lists containing 20 words that were associates of a nonpresented theme word, subjects were scanned during separate test blocks in which they responded separately to true targets (words that had been studied previously), false targets (nonstudied semantic associates of previously studied items), or target controls (nonpresented words that were unrelated to previously studied words). On balance, brain activity during true and false recognition was strikingly similar: compared with a common baseline condition, both true and false recognition

were associated with blood flow increases in various regions that prior studies had shown are commonly activated by memory retrieval tasks, including dorsolateral/anterior prefrontal, medial parietal, and medial temporal regions.

Nonetheless, there was also evidence that true and false recognition could be distinguished: direct comparisons indicated greater activation during true than false recognition in a left temporo-parietal region previously associated with auditory processing and memory. We interpreted this latter finding in light of the aforementioned sensory reactivation hypothesis: Because subjects had heard true targets -- but not false targets -- during the auditory study phase of the experiment, we reasoned that selective left temporo-parietal activation for true recognition might be an auditory sensory signature that reflects memory for auditory/phonological aspects of previously studied words.

One methodological limitation of this early study concerns the fact that in PET imaging experiments, stimuli from different conditions are presented in separate blocks (e.g., all true targets are presented in one block, all false targets are presented in a separate block, and so forth), which is a departure from the typical practice in purely behavioral experiments of intermixing items from different conditions. Soon after we completed the initial PET study, however, *event-related* fMRI methods became available, which allow intermixing of items from different conditions. Schacter, Buckner, Koutstaal, Dale, and Rosen (1997) used event-related fMRI to investigate true and false recognition in a DRM paradigm where true and false targets were randomly intermixed during the recognition test. While the results from this study replicated the PET findings showing that many of the same patterns of brain activity are observed during both true and false recognition, no regions showed greater activation for true than false recognition, including the left temporo-parietal region observed by Schacter et al.

(1996) (for discussion of possible reason why blocked and event-related designs yielded different results, see Johnson, Nolde, Mather, Kounios, Schacter, & Curran, 1997; Schacter et al., 1997).

Our initial neuroimaging studies thus yielded inconclusive evidence regarding the question of whether brain regions involved in sensory/perceptual processing are differentially active during true and false recognition. In a subsequent study that also used event-related fMRI, our lab addressed the issue in a collaborative study with Roberto Cabeza's lab (Cabeza, Rao, Wagner, Mayer, & Schacter, 2001). The logic of this study emerged from previous findings that differences between true and false recognition can be increased when perceptual processing of target materials is increased during encoding (e.g., Schacter, Israel, & Racine, 1999), thereby providing a basis for subjects to differentiate true from false targets during a memory test. If increased perceptual encoding is reflected in patterns of brain activity during retrieval, then true-false differences in brain activity should show a greater contrast.

To produce increased perceptual encoding, prior to scanning Cabeza et al. instructed subjects both to remember lists of semantically associated words and, critically, to try also to remember the source (a man or a woman) that presented the word lists; during encoding, subjects viewed videotapes in which a male source spoke half the words and a female source spoke the other half. The key hypothesis was that on the recognition test, previously studied words – but not semantically associated false targets – would activate regions initially involved in encoding perceptual information related to the sources. Consistent with this general idea, the experiment revealed that the parahippocampal gyrus, a region within the medial temporal lobe that has been linked with processing of contextual information (e.g., Bar & Aminoff, 2003; Bar, Aminoff, & Schacter, 2008), showed greater activation during true than false recognition, perhaps reflecting a lingering effect of contextual encoding processes that occurred for true but not false targets.

Further, there was also greater activity during true as opposed to false recognition in the left parietal cortex (Brodmann's area [BA] 39/40), a region previously implicated in auditory word processing. These data are thus consistent with the early data reported by Schacter et al. (1996) suggesting reactivation of auditory word processing during true but not false recognition.

All of the previous studies used familiar words as target stimuli. In a subsequent study, Slotnick and Schacter (2004) used novel shapes as target stimuli in an attempt to engage more robustly visual processing regions and thereby provide a stronger test of the idea that brain activity associated with the recovery of sensory/perceptual information distinguishes true versus false recognition. All shapes presented in the study list were generated from, and physically related to, prototype shapes that were not presented during encoding. Later, subjects made old/new recognition decisions about previously studied shapes, nonstudied related shapes, and nonstudied unrelated shapes. Previous behavioral studies using such prototype paradigms have shown high levels of false recognition to nonstudied prototype shapes that are perceptually related to previously studied shapes (e.g., Koutstaal, Schacter, Verfaellie, Brenner, & Jackson, 1999; see also, Posner & Keele, 1968). Further, neuroimaging studies of true recognition for pictures had already revealed reactivation during retrieval of some of the same visual processing regions that were active during encoding (Wheeler, Petersen, & Buckner, 2000), and similar sensory reactivation effects were observed during memory for sounds (Nyberg, Habib, McIntosh, & Tulving, 2000; Wheeler et al., 2000). Based on these observations and the sensory reactivation hypothesis discussed earlier, Slotnick and Schacter (2004) hypothesized that true recognition of previously studied shapes, as compared to false recognition of nonstudied related shapes, would be accompanied by a sensory signature involving increased activation of visual processing regions.

Behavioral data revealed that participants made significantly more "old" responses to studied shapes than to related nonstudied shapes (i.e., prototypes); they also made significantly more "old" responses to related nonstudied shapes than to unrelated nonstudied shapes, confirming the presence of a false recognition effect. Consistent with the sensory reactivation hypothesis, analysis of the fMRI data revealed significantly greater activity during true than false recognition in regions within primary visual cortex (e.g., BA 17, 18) that are concerned with processing such features of target stimuli as orientation and color. By contrast, higher-order visual areas in occipito-temporal cortex (e.g., BA 19, 37) showed comparable levels of activity during true and false recognition.

Slotnick and Schacter (2004) also attempted to delineate whether the observed true-false sensory reactivation effects were accompanied by conscious recollection of sensory features of studied shapes, since it is possible that sensory reactivation effects during true recognition reflect nonconscious or implicit priming (Schacter, 1987; Tulving & Schacter, 1990) rather than conscious recollection. Slotnick and Schacter attempted to determine whether activity in visual processing regions is specifically related to conscious memory, which would be indicated by greater activity during "old" than "new" responses to studied items. If, however, brain activity is similar during "old" and "new" responses to studied items, then such activity is likely associated with a nonconscious or implicit form of memory. To address the matter, Slotnick and Schacter compared activation associated with "old" responses to studied shapes (old-hits) and "new" responses to studied shapes (old-misses). This analysis revealed that both old-hits and old-misses were similarly associated with activity in early visual processing regions (BA17, BA18), suggesting that such activity reflects nonconscious memory. By contrast, the old-hits greater than old-misses contrast, which is assumed to reflect conscious memory, revealed activity in late

visual processing regions (BA19, BA37), likely reflecting a form of conscious memory. Thus, activity in early visual processing areas may reflect specific memory for a particular shape, regardless of participants' conscious judgments of whether or not they had seen the item.

This conclusion is consistent with the possibility noted earlier that true and false recognition may be distinguished by nonconscious memory, such as perceptual priming effects that occur only for studied items. Although priming effects in neuroimaging studies are most commonly expressed by *reduced* activity in a particular region during a primed condition compared with an unprimed condition (for reviews, see Henson, 2003; Schacter & Buckner, 1998; Wiggs & Martin, 1998; Schacter, Wig, & Stevens, 2007), this conclusion is based primarily on studies in which familiar words or objects are used as experimental stimuli. In studies that have used unfamiliar materials more akin to the Slotnick and Schacter abstract shapes, such as faces (Henson, Shallice, & Dolan, 2000) or drawings of structurally possible and impossible objects (Schacter, Reiman, et al., 1995), priming-related increases have been observed.

To evaluate the priming hypothesis more directly, Slotnick and Schacter (2006) performed an additional experiment using the identical abstract shape stimuli as in the Slotnick and Schacter (2004) study. However, instead of testing memory with an old-new recognition test as in the aforementioned experiment, Slotnick and Schacter (2006) assessed memory using a priming procedure. During the study phase, participants made line orientation judgments about each shape: they judged whether the lines within each shape were oriented upwards or downwards. Later, during the test phase, subjects made the same judgments about previously presented shapes, physically related shapes that had not been presented earlier, and novel, unrelated shapes. The key finding was that during the test phase, there was increased activity for

old shapes compared with related new shapes in early visual areas (BA 17, 18) but not in late visual areas (BA 19, 37). These findings are consistent with the hypothesis that increased activity in early visual regions during true recognition compared with false recognition reflects a form of priming or nonconscious memory, at least in the visual shape paradigm used by Slotnick and Schacter.

Further evidence bearing on the sensory reactivation hypothesis comes from a study by Kensinger and Schacter (2006; see also, Kensinger & Schacter, 2005a, 2005b) that used fMRI to examine brain activity during accurate and inaccurate retrieval of perceived and imagined items. Adapting a paradigm initially developed by Gonsalves and Paller (2000), prior to scanning participants viewed a series of concrete nouns and formed mental images of the named objects. A picture of the object followed half the names. During the subsequent scan, participants saw the object names and tried to remember whether a corresponding picture had been studied. Kensinger and Schacter found that activity in two regions was associated with accurate assignment of an item to prior pictorial presentation compared with all other conditions: precuneus (BA 7) and left lateral parietal lobe (BA 7/40). Based on previous evidence, Kensinger and Schacter suggested that the precuneus activation could reflect retrieval of sensory details about the perceived picture, whereas the left lateral parietal lobe activation might reflect retrieval of contextual information.

In summary, the neuroimaging studies from our laboratory reviewed so far have provided a good deal of support for the sensory reactivation hypothesis, although the exact regions that have distinguished true from false recognition have varied from study-to-study. Our data are generally consistent with results reported in other laboratories. For example, using a variant of the DRM paradigm discussed earlier, Abe, Okuda, Suzuki, Sasaki, Matsuda, Mori, Tsukuda, and

Fujii (2008) reported greater activity during true than false recognition in the left temporo-parietal regions initially reported by Schacter et al. (1996) and Cabeza et al. (2001). Abe and colleagues interpreted these findings as support for an auditory reactivation effect. In a study by Okado and Stark (2003), subjects studied verbal labels of common objects followed either by a picture of the object or instructions to imagine the object. They were subsequently given a "lie test" in which they were asked to indicate whether or not they had previously seen an actual picture of the object, and were strongly encouraged to tell a lie that they had seen a picture even when they had not. Subjects were then scanned during a standard memory test in which they indicated whether they had actually seen a picture of an object during the study phase. Results revealed that a number of regions showed greater activity during true than false recognition, including bilateral occipital cortices and right parahippocampal gyrus, which Okado and Stark suggested reflects greater recovery of sensory/perceptual information for true versus false memories. The findings concerning parahippocampal gyrus replicate and extend those reported previously by Cabeza et al. (2001).

Although our discussion of brain activity that distinguishes true from false recognition has emphasized the sensory reactivation hypothesis (for evidence from ERP studies that supports the sensory reactivation hypothesis, see Curran, Schacter, Johnson, and Spinks, 2001; Fabiani, Stadler, and Wessels, 2000; Nessler, Mecklinger, & Penney, 2001; Nessler & Mecklinger, 2003; Walla, Endl, Lindinger, Deecke, & Lang, 2000), not all neuroimaging studies of true versus false recognition have focused on this hypothesis. For example, using the DRM paradigm, Kim and Cabeza (2007b) reported that subjective confidence ratings associated with true recognition depend on recollective processes associated with the medial temporal lobe; by contrast,

subjective confidence associated with false recognition seemed to reflect a familiarity-based process associated with activity in fronto-parietal regions.

In a more recent study conducted in our laboratory, Giovanello, Kensinger, Wong, and Schacter (2010) used fMRI to examine brain activity during retrieval of true and false memories in a *memory conjunction error* paradigm. Memory conjunction errors occur when individuals falsely claim to recognize an item because components of the item were contained in previously presented items. For example, after studying compound words such as "blackmail", "jailbird", and "shoestring", participants frequently claim to recognize *conjunction lures* such as "blackbird" in which both parts of the item had been studied previously, and less frequently claim to recognize *feature lures* such as "drawstring" in which one part is studied and one part is novel (e.g., Jones and Jacoby, 2005; Reinitz, Lammers, & Cochran, 1992). Using fMRI to compare brain activity in young and old adults, Giovanello et al. (2010) reported that activity in right anterior hippocampus distinguished between true recognition on the one hand and false alarms to conjunction and feature lures on the other in young but not old adults, likely reflecting a hippocampal contribution to binding and retrieval of components of target items. The older group, by contrast, showed increased activity in right inferior and middle prefrontal cortex during true versus false recognition, extending previous observations of a shift in processing from hippocampal to frontal mechanisms that has been observed in other fMRI studies of aging memory (e.g., Grady, McIntosh, and Craik, 2005; Gutchess, Welsh, Hedden, Bangert, Minear, Liu, and Park, 2005).

Is there a neural signature of false memory retrieval?   In the studies discussed thus far, we have considered brain activity that is preferentially associated with true memory retrieval, focusing in particular on neural evidence that bears on the sensory reactivation hypothesis.

However, it is also important to ask whether there is neural activity that is preferentially associated with false memory retrieval. One point worth noting is that not all forms of false recognition depend on the same neural processes. FMRI evidence in support of this point comes from a study by Garoff-Eaton, Slotnick, and Schacter (2006) that used a variant of the shape prototype paradigm used previously by Slotnick and Schacter (2004). False recognition of new shapes that were perceptually related to previously studied items, like true recognition of the studied shapes, engaged prefrontal, parietal, and medial temporal regions, whereas false recognition of new shapes that were not perceptually related to previously studied shapes engaged distinct temporal regions associated with language processing (perhaps reflecting verbal coding strategies that gave rise to unrelated false recognition).

Although activity in a variety of brain regions has been associated with false memory retrieval in individual studies, attention has focused on regions within the prefrontal cortex that have been linked with retrieval monitoring – that is, evaluating the products of retrieval with respect to task goals. As noted earlier, it is well established that a variety of prefrontal regions show increased activation during standard recognition memory tests. Several of the previously reviewed PET and fMRI studies of false recognition using the DRM and related paradigms have also reported evidence for activation of several prefrontal regions during false recognition, and in some cases, greater activation during false than true recognition. Schacter et al. (1996) reported that a region in the dorsolateral/anterior prefrontal cortex, that has been associated with retrieval monitoring (e.g., Dobbins, Foley, Schacter, & Wagner, 2002; Dobbins, Rice, Wagner, & Schacter, 2003; Rugg, Fletcher, Frith, Frackowiak, & Dolan, 1996), showed greater activity during false than true recognition, perhaps reflecting the need for evaluation or monitoring of the strong sense of familiarity produced by false targets. Consistent with this suggestion, analyses of

event-related time courses indicated a delayed onset for anterior prefrontal activity during both true and false recognition compared with other brain regions. Although various interpretations of this observation are possible (see Schacter et al., 1997), it is consistent with the idea that anterior prefrontal activity (especially on the right) reflects a late-occurring evaluation or monitoring of the products of retrieval. Cabeza et al. (2001) and Slotnick and Schacter (2004) provided additional evidence for greater activation of right prefrontal cortex during false than true recognition (see also, Treyer, Buck, & Schnider, 2003; von Zerssen, Mecklinger, Opitz, & von Cramon, 2001; for converging evidence from ERP studies, see Curran et al., 2001, Fabiani et al., 2000, and Goldmann et al., 2003; but see Düzel et al., 1997, Nessler et al., 2001, and Nessler & Mecklinger, 2003, for contrasting results).

Although much attention has focused on the possibility that increased frontal lobe activity during false versus true recognition is associated with retrieval monitoring operations, researchers have also considered other factors. One idea is that activity in certain frontal lobe regions reflects the influence of conceptual or semantic processing that contributes to false recognition. For example, Garoff-Eaton, Kensinger, and Schacter (2007) compared conceptual false recognition – false alarms that result from semantic or associative similarities between studied and tested items – with perceptual false recognition – false alarms that result from physical similarities between studied and tested items. Garoff-Eaton et al. found that multiple regions within the prefrontal cortex (BAs 6, 8, 9, 44, 45, 46, 47) showed increased activity during conceptual false recognition compared with true recognition, but not during perceptual false recognition compared with true recognition. Garoff-Eaton et al. considered it unlikely that these findings are attributable to post-retrieval monitoring, because such monitoring should be required during both conceptual and perceptual false

recognition when subjects attempt to evaluate the accuracy of retrieved information. Instead, Garoff-Eaton et al. noted that regions within the frontal cortex (especially left inferior frontal cortex) show increased activity in a variety of conditions that involved conceptual processing, including semantic elaboration (Demb, Desmond, Wagner, Vaidya, Glover, & Gabrieli 1995; Kirchhoff, Shapiro, & Buckner, 2005), word generation (Petersen, Fox, Posner, Mintun, & Raichle, 1988), recovery of meaning/semantic retrieval (Wagner, Paré-Blagoev, Clark, & Poldrack, 2001), and semantic (versus perceptual) relational memory (Prince, Daselaar, & Cabeza, 2005). It therefore seems reasonable to hypothesize that the increased activity in left inferior and perhaps other frontal regions during conceptual versus perceptual false recognition reflects the retrieval of conceptual information about the meaning or gist of what was studied, which drives the false recognition response. Paz-Alonso, Ghetti, Donohue, Goodman, and Bunge (2008) offered a similar interpretation of left inferior frontal responses in a study that examined false recognition using the DRM paradigm.

Additional relevant evidence comes from work by Dennis, Kim, and Cabeza (2007, 2008) concerning age differences in true and false recognition, using an adapted version of the DRM semantic associates paradigm in which they used fMRI to measure brain activity during encoding (Dennis et al., 2007) or retrieval (Dennis et al., 2008). Both studies provided evidence for age-related increases in left middle temporal gyrus during processing of false memories. In light of previous evidence linking the left middle temporal region with processing of semantic information (e.g., Wise & Price, 2006), Dennis et al. suggest that their findings may reflect increased reliance on semantic gist information in older adults, an idea that is consistent with other behavioral evidence documenting age-

related increases in gist-based processing (e.g., Koutstaal & Schacter, 1997; Norman & Schacter, 1997; Tun, Wingfield, Rosen, & Blanchard, 1998; see Paz-Alonso et al., 2008, for discussion of related issues in children and adolescents).

Neuroimaging of true and imaginary memories

Memory researchers have long been interested in the relationship between imagination and memory distortion, focusing mainly on the ways in which imagining events can contribute to the development of false memories for those events (e.g., Garry, Manning, Loftus, & Sherman, 1996; Gonsalves & Paller, 2000; Johnson & Raye, 1981; Goff & Roediger, 1998; Kensinger & Schacter, 2006; Loftus, 2003). During the past several years, however, there has been increasing interest in how memory contributes to imagination, as reflected in a recent outpouring of studies concerning the role of memory in imagining or simulating possible future events, also referred to as episodic future thought (for reviews, see Buckner & Carroll, 2007; Schacter, Addis, & Buckner, 2007, 2008; Szpunar, 2010). While some of the issues considered in the emerging literature concerning these "imaginary memories" focus on the nature of prospective cognition and thus are not directly relevant to the current discussion, a number of neuroimaging studies reported recently have examined the extent to which neural activity associated with remembering the past can be distinguished from that associated with imagining the future, and those studies will constitute the focus of our discussion (for an analysis of how cognitive characteristics of episodic future thought can help to distinguish between true and false intentions, see Granhag & Knieps, in press).

In the first study from our laboratory to examine the issue, Addis, Wong, and Schacter (2007) scanned participants while they were either remembering a past experience or

imagining an event that might occur in the future. Addis et al. divided each of these tasks into two phases. In the initial *construction* phase, participants generated a remembered or imagined event in response to a cue (e.g., "dress") and made a button-press when they had an event in mind, which typically required about seven or eight seconds. In the immediately following *elaboration* phase, participants generated as much detail as possible about the remembered or imagined event. The most striking finding was that brain activity was highly similar during remembering the past and imagining the future. This overlap was most apparent during the elaboration phase, when participants focused on generating details about the remembered or imagined event. A network of brain regions that had previously been implicated in the retrieval of episodic memories (Maguire, 2001) showed common activation during both remembering and imagining, including the hippocampus, parahippocampal and retrosplenial cortices, medial prefrontal and frontopolar cortices, and lateral parietal lobe.

The common activation observed in the hippocampus, a structure long known to be critically involved in aspects of memory, was especially intriguing, possibly reflecting the retrieval, integration, or encoding of event details into the remembered or imagined representation, and complementing evidence reported by Hassabis, Kumaran, Vann, and Maguire (2007) that amnesic patients with hippocampal damage, who have great difficulty remembering past events, are also impaired when they are asked to imagine novel scenes (see also Andelman, Hoofien, Goldberg, Aizenstein, & Neufeld, 2010, and Kwan, Carson, Addis, & Rosenbaum, 2010, for similar observations; but see Maguire, Vargha-Khadem, & Hassabis, 2010, and Squire, van der Horst, McDuff, Frascino, Hopkins, & Mauldin, 2010, for evidence that not all hippocampal amnesics have problems imagining future experiences or novel scenes).

The construction phase was also associated with common activity during remembering and imagining. Critically important for the present discussion, however, this phase also revealed some neural differences. Most interestingly, the right hippocampus was engaged to a greater extent when participants imagined future events than when they remembered past events. Because the hippocampus has been implicated in relational processing (i.e., linking together previously unrelated items; Eichenbaum & Cohen, 2001), Addis et al. suggested that this finding might reflect the additional relational processing required when one must recombine disparate details into an imagined future event (for further discussion, see Schacter & Addis, 2007, 2009).

A related observation was that in all regions that did exhibit significant differences, imagined events were associated with greater activity than remembered events, a finding that was also reported in similar studies by Okuda et al. (2003) and Szpunar, Watson, and McDermott (2007). Addis et al. suggested that this pattern reflects the more intensive constructive processes required by imagining future events relative to retrieving past events. Even though both the remembering and imagining tasks involve retrieval of information from memory and engage common memory networks, only the imagining task requires that details from various past events are flexibly recombined into a novel future event and that this imagined event is plausible given one's intentions for the future. Accordingly, additional regions supporting these processes are recruited by the future event task.

Following up on the foregoing findings with respect to hippocampal activity, Addis and Schacter (2008) examined the relationship between brain activity and the amount of detail reported for remembered and imagined events during the elaboration phase. Addis and Schacter observed that activity in the left *posterior* hippocampus was correlated with the amount of detail comprising both remembered and imagined events, whereas the left *anterior* hippocampus

responded specifically to the amount of detail comprising imagined but not remembered events. In line with the previous discussion, Addis and Schacter suggested that this latter finding could reflect activity associated with the recombining of details into an imagined future event.

Further examining the possibility that hippocampal activity distinguishes remembering the past from imagining the future, Addis, Cheng, Roberts, and Schacter (in press) asked participants to remember specific past events or imagine specific future events, as in the aforementioned work, but in addition asked participants to remember general, routine events (e.g., having brunch after attending church) or to imagine general events that might occur in their personal futures (e.g., reading the newspaper each morning). The logic here is that a region that is responsive to the amount of detail recombined into a coherent imagined episode should show more activity when constructing specific future events relative to general future events (as well as specific and generic past events). Focusing on the increased right hippocampal activity previously associated with constructing imagined future events, Addis et al. (in press) found that this activity was evident only for specific imagined events; there was no evidence for right hippocampal activity during construction of generic imagined events. Thus, the results appear to provide evidence that right hippocampal activation constitutes a neural signature associated with the construction and encoding of specific imagined events. Addis et al. (in press) discuss several possible theoretical accounts of this finding.

An additional study from our laboratory has provided further information concerning neural activity that distinguished remembered from imagined experiences by using an *experimental recombination paradigm* (Addis, Pan, Vu, Laiser, & Schacter, 2009). Participants initially provided episodic memories of actual experiences that included details about a *person, object, and place* involved in that event. During a later scanning session, they were cued to recall

some of the events that had actually occurred. Critically, for the conditions in which they imagined events, the experimenters randomly recombined details concerning person, object, and place from separate episodes that they had recalled previously. During scanning, participants were given cues for a person, object, and place taken from distinct episodes, and were instructed to imagine a single, novel episode that included the specified details. In some cases, participants were instructed to imagine possible future events, whereas in others, they were instructed to imagine events that might have occurred in the past. Brain activity during remembering and imagining recruited a network of regions similar to that observed in previous research, including medial temporal, parietal and prefrontal structures as well as some posterior visual regions (lingual and fusiform gyri), thereby providing additional evidence for shared processes during remembering and imagining. However, Addis et al. (2009) also reported evidence that distinct subsystems of this common network were preferentially associated with imagining and remembering, respectively. The *imagining network* consisted of medial temporal lobe including anterior hippocampus, bilateral medial prefrontal cortex, inferior frontal gyrus, polar and posterior temporal cortex, and medial parietal cortex. The remembering network included posterior visual cortices such as fusiform, lingual and occipital gyri and cuneus, as well as parahippocampal gyrus and posterior hippocampus. Addis et al. (2009) interpreted the finding that visual cortices were preferentially associated with the remembering network in light of the sensory reactivation hypothesis we discussed earlier in relation to distinguishing between true and false memories, i.e., that reactivation of sensory-perceptual and contextual details during retrieval recruits the neural regions involved in the original processing of the remembered information.

The foregoing studies, as well as related studies from other laboratories (e.g., Andrews-Hanna, Reidler, Sepulcre, Poulin, & Buckner, 2010; Botzung, Denkova, & Manning, 2008; Hassabis, Kumaran, & Maguire, 2007; Spreng & Grady, 2010; Szpunar, Chan, & McDermott, 2009; Weiler, Suchan, & Daum, 2010), converge with studies of true vs. false memories in two ways. First, both sets of studies reveal extensive overlap in the brain activity that accompanies remembering of actual experiences compared with either imaginary or false memories. Second, both sets of studies show that despite the overlap, the neural activity associated with true memories can be distinguished from that associated with imaginary or false memories, with some evidence from both lines of work supporting the sensory reactivation hypothesis.

Concluding Comments

Our review indicates that research that has used neuroimaging methods to distinguish among true, false, and imaginary memories has progressed over the past decade. Some reliable findings have emerged, and several studies suggest that despite striking similarities in the brain activity that accompanies true, false, and imaginary memories, conditions exist in which neuroimaging techniques can distinguish patterns of brain activity associated with them. However, it is clear that this research is in a nascent stage, and that many obstacles need to be overcome before it will be possible to seriously consider applying neuroimaging technology to the courtroom cases in which the veracity of memory is at stake. We will first consider several key limitations of the studies discussed in this chapter, and then broaden the discussion to address concerns regarding potential applications to the courtroom.

Limitations of current research   One limitation of current research is that much of what we know about neuroimaging of true, false, and imaginary memories comes from studies on healthy young adults, primarily college students. Although we reviewed several fMRI studies

of true versus false memories involving older adults (Dennis et al., 2007, 2008; Giovanello et al., 2010) and young children (Paz-Alonso et al., 2008), it seems fair to state that there has been a relative lack of subject diversity in published studies, thereby leaving unknown the effect of such variables as education, intelligence, socio-economic status, psychiatric disorders, and medications on our ability to identify a memory's veracity (cf., Greely & Illes, 2007). Because diverse populations are encountered in the courtroom, it seems clear that future studies will need to examine more diverse populations.

A second important limitation concerns the kinds of materials that are typically used in neuroimaging studies. Many of the studies we reviewed examined memory for such simple materials as words, shapes, and faces. Using these kinds of simple materials allows researchers to exercise precise experimental control over what is encoded and recalled, which in turn facilitates attempts to specify the neural and cognitive processes that support memory. However, using such simple materials also constrains the generalizability of much laboratory research to the courtroom, where disputes concerning the veracity of memory often involve more complex and richer autobiographical memories. Underscoring the need for caution, there is evidence showing different patterns of activation for everyday autobiographical memories and memories based on information encoded in the laboratory (Cabeza et al., 2004; McDermott, Szpunar, & Christ, 2009).

Note, however, that the recent research that we reviewed concerning neural differences between remembering the past and imagining the future has relied on remembering and imagining of rich everyday experiences. But such an approach is not easy to apply to studying true versus false memories: while it is relatively straightforward to ask participants to imagine experiences that might occur in their personal futures, it is more difficult to create subjectively

compelling false memories for actual autobiographical events. Though there are empirical

demonstrations that compelling false autobiographical memories can in fact be created (e.g.,

Hyman & Billings, 1998; Loftus & Pickrell, 1995; Mazzoni & Memon, 2003), such memories

are often observed in only a minority of participants. Moreover, they typically involve only a

single experience, whereas fMRI studies of true versus false memories have used experimental

paradigms in which multiple items contribute to both kinds of memories.

This latter point highlights a third limitation of the studies reviewed here: researchers

have drawn conclusions about true, false, and imaginary memories by averaging across subjects

and events (whether those events are words in a list or actual autobiographical experiences).

Neuroimaging techniques have an inherently low signal to noise ratio, thereby hindering the

ability to detect meaningful patterns of activation on any given trial. Accordingly, researchers

have used statistical techniques that increase power by averaging effects across multiple trials

and subjects. In the courtroom, where only a single event or possibly a few events are of interest,

neuroimaging techniques currently lack the necessary power to be useful.

However, recent methodological advances show promise for detecting a participant's

subjective experience of remembering during a single trial. For example, Rissman, Greely, and

Wagner (2010) recently used a classification technique known as multivoxel pattern analysis to

determine with a high degree of accuracy when individual participants did and did not *believe*

that they were remembering a single event (see also, Chadwick, Hassabis, Weiskopf, & Maguire,

2010). By contrast, they reported that the *objective status* of memory for single events could not

be successfully decoded by the pattern classifier. Although classification of the subjective state

of an individual is a significant achievement that has potential applications in the courtroom

(e.g., in assessing perjury), ideally a classificatory technique should reliably establish the

objective status of a memory as well, so that the technique could be applied to such issues as determining whether an individual was present at a crime scene. Further experiments are needed to determine whether this goal can be achieved. At the present time, we find ourselves in agreement with Bernstein and Loftus (2009), who concluded based on their review of cognitive and neuroimaging studies of true versus false memories that "…it might be virtually impossible to tell reliably if a particular memory is true or false without independent corroboration (p.373)."

Even if each of the three preceding caveats could be overcome, a fourth limitation would have to be addressed: ensuring that individuals cannot be instructed to use strategies that allow them to "beat" a test that can reliably distinguish true from false or imaginary memories. For example, a recent behavioral experiment presented a new method of distinguishing true from false past autobiographical events using an autobiographical version of the Implicit Association Test (aIAT; Sartori, Agosta, & Zogmaister, 2008). The standard IAT has been used in many studies to reveal potentially unconscious or hidden biases by asking participants to pair two concepts (e.g., black/good) under time pressure (e.g., Greenwald, Poehlman, Uhlmann, & Banaji, 2009; https://implicit.harvard.edu). The easier the pairing of two concepts in one's mind, the faster the response should be, thus reflecting the strength of association between the two concepts and revealing one's implicit attitude toward a concept. Sartori and colleagues (2008) made use of this method by presenting respondents with categorization trials pairing true autobiographical events with either true or false autobiographical events. Most of the autobiographical events were related to criminal activities, such as stealing a CD, using cocaine, or driving drunk. The researchers found that associating true autobiographical events with other true autobiographical events sped up response times significantly compared to pairing true with false autobiographical events, thus providing a means to detect which autobiographical events

were true and which were false. The authors claimed to be able to classify 91% of participants correctly as guilty or innocent using the aIAT, which suggested that the test could be a useful tool in forensic settings.

However, only a year later, Verschuere, Prate, and De Houwer (2009) reported ways in which respondents could cheat the aIAT to appear innocent when they were in fact guilty. Participants who were instructed to slow down their responses for those trials that paired a confessing statement (true autobiographical event) with a true statement were classified as innocent and could not be identified as cheaters. Participants with prior experience with the aIAT were even better at beating the test, but prior experience was not necessary for successful, undetectable faking. The aIAT can be cheated if a respondent is provided with simple instructions, lessening its potential use for distinguishing true and false autobiographical events in forensic settings unless measures can be developed that allow fakers to be detected reliably. Agosta, Ghirardi, Zogmaister, Castiello, & Sartori (in press) have recently provided evidence that they can detect faking on the aIAT, but further research is necessary before the technique can be applied to real-word settings. Similarly, strategies that allow participants to beat the test have also been reported for the polygraph (Ben-Shakhar & Elaad, 2003), and it has been suggested (Sartori et al., 2008) that fMRI-based lie detection relying on frontal lobe activity during deception might be thwarted by the use of conscious strategies that are known to activate relevant frontal structures (e.g., Cole & Schneider, 2007). Therefore, any neuroimaging-based method to distinguish between true and false or imaginary memories will have to address and overcome attempts to "beat the test" if that method is going to be useful in everyday settings. Because we are unaware of any tests that meet this standard, these considerations reinforce our

earlier point that current neuroimaging-based approaches to distinguishing among true, false, and imaginary memories are not yet ready for real-world application.

Applications to the courtroom   During the last decade, there has been considerable debate over the legality of admitting neuroimaging results as evidence into the courtroom. While the courts have not yet specifically addressed research on true, false, and imaginary memories, related work on lie detection has received considerable attention. Because some of the issues that arise when discussing the possible use of neuroimaging to detect deception are similar to those that arise when attempting to detect the veracity of memories, the treatment of fMRI data for lie detection in the courts likely presages the challenges that neuroimaging of false memories will face.

There are several impediments to fMRI-based lie detection techniques gaining admissibility in the courts. The Daubert standard (Daubert vs. Merrell Dow Pharmaceuticals, 2003), the benchmark test of the admissibility of scientific evidence in most federal courts, specifies four criteria that should be weighed when considering the admissibility of a scientific theory or technique: it should be empirically testable, subjected to peer review and publication, have a known and acceptably low error rate, and be generally accepted within the relevant scientific community.

Most of the criticism of fMRI-based lie detection's admissibility has centered on the latter two criteria, including attention to limitations such as those we just considered with respect to memory studies. In an analysis of the twenty-eight peer reviewed publications examining deception versus truth telling in neuroimaging studies, Wagner (2010, p. 22) concludes that " the published literature reveals no data that provides unambiguous evidence regarding the sensitivity and specificity of fMRI-based neuroscience methods in the detection of lies at the individual-

subject or the individual-event levels." Additional objections to the courtroom admissibility of fMRI-based lie detection include lack of replication (Greely & Illes, 2007) and real-world applicability (Kanwisher, 2009), individual differences in brain function (Raichle, 2010), and inconsistencies in reported areas of brain activation (Alexander, 2007).

Even if in the future fMRI-based lie detection were to satisfy the guidelines outlined by the Daubert standard, judges could still deem it inadmissible if they determine that the probative value of the evidence is outweighed by its potential to confuse or mislead the jury. Additionally, critics have argued that it violates an individual's fourth amendment rights against unreasonable search and seizure (Luber, Fisher, Appelbaum, Ploesser, & Lisanby, 2009; New, 2008) and fifth amendment rights against self-incrimination (Holloway, 2008; Luber et. al., 2009; New, 2008).

Just this year, attorneys introduced fMRI-based lie detection in a U.S. federal court for the first time, prompting a precedent-setting Daubert hearing (Miller, 2010). Cephos, one of two companies specializing in fMRI-based lie detection, was hired to perform the testing, and scanned the defendant while he responded both truthfully and dishonestly to questions unrelated to the case in order to determine a baseline neural response for deceptive responses. The defendant was then scanned while being asked questions involving details of the case to compare neural activity across the two conditions. The judge decided that the evidence was inadmissible, finding that while fMRI-based lie detection is testable and has been subjected to peer-review, it lacks established error rates and is not generally accepted by the scientific community. Additionally, he cited that the reported error rates came from controlled laboratory experiments and could not be directly applied to real-world scenarios.

Given this precedent, the current status of fMRI-based lie detection in the courts portends the need for considerable advancement and refinement in the basic science underlying

neuroimaging of true, false, and imaginary memories before this research can have a place in the courtroom. While we are hopeful that the field will make significant advances in the years to come, we believe that a cautionary stance is currently necessary in light of the present state of the art.

References

Abe, N. (2009). The neurobiology of deception: evidence from neuroimaging and loss-of-function studies. *Current Opinion in Neurology, 22*, 594-600.

Abe, N., Okuda, J., Suzuki, M., Sasaki, H., Matauda, T., Mori, E., Tsukada, M., & Fujii, T. (2008). Neural correlates of true memory, false memory, and deception. *Cerebral Cortex, 18*, 2811-2819.

Addis, D.R., Cheng, T., Roberts, R., & Schacter, D.L. (in press). Hippocampal contributions to the episodic simulation of specific and general future events. *Hippocampus*. DOI: 10.1002/hipo.20870

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

Addis, D. R., & Schacter, D. L. (2008). Constructive episodic simulation: Temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus, 18*, 227-237.

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

Agosta, S., Ghirardi, V., Zogmaister, C., Castiello, U., & Sartori, G. (in press). Detecting fakes of the *autobiographical* IAT. *Applied Cognitive Psychology*. DOI: 10.1002/acp.1691

Alexander, A. (2007). Functional magnetic resonance imaging lie detection: Is a "brainstorm" heading toward the "gatekeeper?" *Houston Journal of Health Law and Policy*, 7, 1-56.

Aminoff, E., Schacter, D.L., & Bar, M. (2008). The cortical underpinnings of context-based memory distortion. *Journal of Cognitive Neuroscience, 20*, 2226-2237.

Andelman, F., Hoofien, D., Goldberg, I., Aizenstein, O., & Neufeld, M.F. (2010). Bilateral hippocampal lesion and a selective impairment of the ability for mental time travel. *Neurocase, 16*, 426-435.

Andrews-Hanna, J., Reidler, J., Sepulcre, J., Poulin, R., & Buckner, R. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron 65*, 550-562.

Bar, M., and Aminoff, E. (2003). Cortical analysis of visual context. *Neuron 38*, 347-358.

Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience, 28*, 8539-8544.

Ben-Shakher, G. & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology 88*, 131-151.

Bernstein, D.M. & Loftus, E.F. (2009). How to tell if a particular memory is true or false. *Perspectives on Psychological Science, 4*, 370-374.

Botzung, A., Denkova, E., & Manning, L. (2008). Experiencing past and future personal events: Functional neuroimaging evidence on the neural bases of mental time travel. *Brain and Cognition, 66*, 202-212.

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

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

Cabeza, R., Prince, S.E., Daselaar, S.M., Greenberg, D.L., Budde, M., Dolcos, F., LaBar, K.S., Rubin, D.C. (2004). Brain activity during episodic retrieval of autobiographical and laboratory events: An fMRI study using a novel photo paradigm. *Journal of Cognitive Neuroscience*, *16*, 1583-1594.

Chadwick, M.J., Hassabis, D., Weiskopf, N., Maguire, E.A. (2010). Decoding individual episodic memory traces in the human hippocampus. *Current Biology*, *20*, 544-547.

Clarke, R.A. (2004). *Against all enemies: Inside America's war on terror*. New York: Free Press.

Cole, M.W. & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage 37*, 343-360.

Curran, T., Schacter, D. L., Johnson, M. K., & Spinks, R. (2001). Brain potentials reflect behavioral differences in true and false recognition. *Journal of Cognitive Neuroscience, 13*, 201-216.

Cutler, B.L. & Penrod, S.D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.

Daubert v. Merrell Dow Pharmaceuticals. 509 U.S. (1993).

Demb, J.B., Desmond, J.E., Wagner, A.D., Vaidya, C.J., Glover, G.H., and Gabrieli, J.D.

(1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: a functional MRI study of task difficulty and process specificity. *Journal of Neuroscience, 15,* 5870-5878.

Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation: an fMRI study. *Neuropsychologia, 45,* 3157-3166.

Dennis, N. A., Kim, H., & Cabeza, R. (2008). Age-related differences in brain activity during true and false memory retrieval. *Journal of Cognitive Neuroscience, 20,* 1390-1402.

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

Dobbins, I.G., Foley, H., Schacter, D.L., and Wagner, A.D. (2002). Executive control during episodic retrieval: multiple prefrontal processes subserve source memory. *Neuron, 35,* 989-996.

Dobbins, I.G., Rice, H.J., Wagner, A.D., and Schacter, D.L. (2003). Memory orientation and success: separable neurocognitive components underlying episodic recognition. *Neuropsychologia, 41,* 318-333.

Düzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H. J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences, 94,* 59731-59738.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain.* New York: Oxford University Press.

Fabiani, M., Stadler, M. A., & Wessels, P. M. (2000). True but not false memories produce a sensory signature in human lateralized brain potentials. *Journal of Cognitive Neuroscience, 12,* 941-949.

Gallo, D.A. (2006). *Associative illusions of memory.* New York: Psychology Press.

Garoff, R. J., Slotnick, S. D., & Schacter, D. L. (2005). The neural origins of specific and general memory: The role of the fusiform cortex. *Neuropsychologia, 43*, 847-859.

Garoff-Eaton, R.J., Slotnick, S.D., & Schacter, D.L. (2006). Not all false memories are created equal: The neural basis of false recognition. *Cerebral Cortex, 16,* 1645-1652.

Garoff-Eaton, R.J., Kensinger, E.A., & Schacter, D.L. (2007). The neural correlates of conceptual and perceptual false recognition. *Learning and Memory, 14*, 684-692.

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

Giovanello, K.S., Kensinger, E.A., Wong, A.T., & Schacter, D.L. (2010). Age-related neural changes during memory conjunction errors. *Journal of Cognitive Neuroscience, 22,* 1348-1361.

Goff, L. M., & Roediger, H. L., III (1998). Imagination inflation for action events- Repeated imaginings lead to illusory recollections. *Memory & Cognition, 26*, 20-33.

Goldmann, R.E., Sullivan, A.L., Droller, D.B.J., Rugg, M.D., Curran, T., Holcomb, P.J., Schacter, D.L., Daffner, K.R., and Budson, A.E. (2003). Late frontal brain potentials distinguish true and false recognition. *Neuroreport, 14*, 1717-1720.

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

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

Grady, C. McIntosh, A.R., and Craik, F.I.M. (2005). Task-related activity in prefrontal cortex and its relation to recognition memory performance in young and old adults. *Neuropsychologia*, *43*, 1466-1481.

Granhag, P.A. & Knieps, M. (in press). Episodic future thought: illuminating the trademarks of forming true and false intentions. *Applied Cognitive Psychology*. DOI: 10.1002/acp.1674

Greely, H.T. & Illes, J. (2007). Neuroscience-based lie detection: the urgent need for regulation. *American Journal of Law and Medicine*, *33*, 377-431.

Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., & Banaji, M.R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology 97*, 17-41.

Gutchess, A., Welsh, R.C., Hedden, T., Bangert, A., Minear, M., Liu, L.L., Park, D.C. (2005). Aging and the neural correlates of successful picture encoding: Frontal activations compensate for decreased-medial temporal lobe activity. *Journal of Cognitive Neuroscience, 17*, 84-96.

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience, 27*, 14365-14374.

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

Henson, R. N. (2003). Neuroimaging studies of priming. *Progress in Neurobiology, 70*, 53-81.

Henson, R. N., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable

forms of repetition priming. *Science, 287*, 1269-1272.

Holloway, M. B. (2008). One image, one thousand incriminating words: images of brain activity and the privilege against self-incrimination. *Temple Journal of Science, Technology & Environmental Law*, 27, 141-175.

Hyman, I.E. & Billings, J. (1998). Individual differences and the creation of false childhood memories. *Memory, 6*, 1-20.

Jacobs, W.J. & Nadel, L. (1998). Neurobiology of reconstructed memory. *Psychology, Public Policy, and the Law, 4*, 1110-1134.

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General, 117*, 371-376.

Johnson, M. K., Nolde, S. F., Mather, M., Kounios, J., Schacter, D. L., & Curran, T. (1997). The similarity of brain activity associated with true and false recognition memory depends on test format. *Psychological Science, 8*, 250-257.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.

Jones, T.C. and Jacoby, L.L. (2005). Conjunction errors in recognition memory: Modality free-errors for older but not for you adults. *Acta Psychologica, 120*, 55-73.

Kanwisher, N. (2009). The use of fMRI in lie detection: what has been shown and what has not. In Bizzi, E. & Hyman, S.E. (eds), *Using imaging to identify deceit: Scientific and ethical questions* (pp. 7–13). Cambridge, MA: American Academy of Arts and Sciences.

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

Kensinger, E.A. & Schacter, D.L. (2005b). Retrieving accurate and distorted memories: Neuroimaging evidence for the effects of emotion. *NeuroImage, 27,* 167-177.

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

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

Kim, H. & Cabeza, R. (2007b). Trusting our memories: Dissociating the neural correlates of confidence in veridical vs. illusory memories. *Journal of Neuroscience, 27*, 12190-12197.

Kirchhoff, B.A., Schapiro, M.L., and Buckner, R.L. 2005. Orthographic distinctiveness and semantic elaboration provide separate contributions to memory. *Journal of Cognitive Neuroscience, 17*, 1841-1854.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37*, 555-583.

Koutstaal, W., Schacter, D. L., Verfaellie, M., Brenner, C. J., & Jackson, E. M. (1999). Perceptually based false recognition of novel objects in amnesia: Effects of category size and similarity to category prototypes. *Cognitive Neuropsychology, 16*, 317-341.

Kwan, D., Carson, N., Addis, D.R., & Rosenbaum, R.S. (2010). Deficits in past remembering extend to future imagining in a case of developmental amnesia. Neuropsychologia, 48, 3179-3186.

Loftus, E.F. (1979). *Eyewitness testimony*. Cambridge: Harvard University Press.

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

Loftus, E. F. (2003). Make-believe memories. *American Psychologist, 58*, 867-873.

Loftus, E.F. & Davis, D. (2006). Recovered memories. *Annual Review of Clinical Psychology, 2*, 469-498.

Loftus, E.F. & Ketcham, K. (1994). *The myth of repressed memory*. New York: St. Martin's Press.

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

Luber B., Fisher C., Appelbaum P.S., Ploesser, M., & Lisanby, S.H. (2009). Non-invasive brain stimulation in the detection of deception: scientific challenges and ethical consequences. *Behavioral Sciences and the Law*, 27, 191-208.

Maguire, E. A. (2001). Neuroimaging studies of autobiographical event memory. *Philosophical Transactions of the Royal Society of London B, 356*, 1441-1451.

Marche, T.A., Brainerd, C.J., & Reyna, V.F. (2010). Distinguishing true from false memories in forensic contexts: Can phenomenology tell us what is real? *Applied Cognitive Psychology*, *24*, 1168-1182. DOI: 10.1002/acp.1629

Maguire, E.A., Vargha-Khadem, F., & Hassabis, D. (2010). Imagining fictitious and future experiences: Evidence from developmental amnesia. *Neuropsychologia, 48*, 3187-3192.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory and Cognition, 25*, 826-837.

Mazzoni, G. & Memon, A. (2003). Imagination can create false autobiographical memories. *Psychological Science, 14*, 186-188.

McDermott, K.B., Szpunar, K.K., Christ, S.E. (2009). Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, *47*, 2290-2298.

McNally, R.J. & Geraerts, E. (2009). A new solution to the recovered memories debate. *Perspectives on Psychological Science, 4*, 126-134.

Miller, G. (2010). FMRI lie detection fails a legal test. *Science*, *328*, 1336-7.

Nessler, D., and Mecklinger, A. (2003). ERP correlates of true and false recognition after different retention delays: stimulus- and response-related processes. *Psychophysiology, 40,* 146-159.

Nessler, D., Mecklinger, A., & Penney, T. B. (2001). Event related brain potentials and illusory memories: the effects of differential encoding. *Cognitive Brain Research, 10*, 283-301.

New, J.G. (2008). If you could read my mind: Implications of neurological evidence for twenty-first century criminal jurisprudence. *The Journal of Legal Medicine*, 29, 179-198.

Norman, K. A., & Schacter, D. L. (1997). False recognition in young and older adults: Exploring the characteristics of illusory memories. *Memory and Cognition*, *25*, 838-848.

Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Acadame of Science of the United States of America, 97*, 11120-11124.

O'Connell, S. (2008). The perils of relying on memory in court. http://www.telegraph.co.uk/technology/3778272/The-perils-of-relying-on-memory-in-court.html.

Okado, Y., & Stark, C. (2003). Neural processing associated with true and false memory retrieval. *Cognitive, Affective, and Behavioral Neuroscience, 3*, 323-334.

Okado, Y. & Stark, C. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning and Memory, 12*, 3-11.

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

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*, 2208-2216.

Pendergrast, M. (1995). *Victims of memory*. Hinesburg, VT: Upper Access.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., and Raichle, M.E. 1988. Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331,* 585-589.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Prince, S.E., Daselaar, S.M., and Cabeza, R. 2005. Neural correlates of relational memory: successful encoding and retrieval of semantic and perceptual associations. *Journal of Neuroscience, 25,* 1203-1210.

Raichle, M. (2010). What is an fMRI? In M. S. Gazzaniga & J. S. Rakoff (Eds.), *A judge's guide to neuroscience: A concise introduction* (pp. 5-12). Santa Barbara: University of California, Santa Barbara.

Reinitz, M.T., Lammers, W.J., and Cochran, B.P. (1992). Memory-conjunction errors:

Miscombinations of stored stimulus features can produce illusions of memory. *Memory & Cognition, 20,* 1-11.

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*, 9849-9854.

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

Rugg, M.D., Fletcher, P.C., Frith, C.D., Frackowiak, R.S.J., and Dolan, R.J. (1996). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain, 119*, 2073-2083.

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S.D., & Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological Science 19*, 772-780.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501-518.

Schacter, D.L. (1996). *Searching for memory: The brain, the mind, and the past*. New York: Basic Books.

Schacter, D.L. (2001). *The seven sins of memory: How the mind forgets and remembers*. New York and Boston: Houghton-Mifflin.

Schacter, D.L. (2004). The fog of war. *The New York Times*, April 5, A25.

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

Schacter, D. L., & Addis, D. R. (2009). On the nature of medial temporal lobe contributions to the constructive simulation of future events. *Philosophical Transactions of the Royal Society of London B, 364*, 1245-1253.

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., Addis, D. R., & Buckner, R. L. (2008). Episodic simulation of future events: Concepts, data, and applications. *The Year in Cognitive Neuroscience, Annals of the New York Academy of Sciences, 1124*, 39-60

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron, 20*, 185-195.

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

Schacter, D. L., Israel, L., & Racine, C. A. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1-24.

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., et al. (1996a). Neuroanatomical correlates of veridical and illusory recognition memory: Evidence from positron emission tomography. *Neuron, 17*, 267-274.

Schacter, D. L., Reiman, E., Uecker, A., Polster, M. R., Yun, L. S., & Cooper, L. A. (1995). Brain regions associated with retrieval of structurally coherent visual information. *Nature, 376*, 587-590.

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

Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology, 17*, 171-176.

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence*. New York: Doubleday.

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12*, 171-181.

Semmler, C., & Brewer, N. (2010). Eyewitness memory. In J. Brown & E. Campbell (Eds.), *The Cambridge handbook of forensic psychology* (pp. 49-57). Cambridge, UK: Cambridge University Press.

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

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

Spence, S.A. & Kaylor-Hughes, C.J. (2008). Looking for truth and finding lies: the prospects for a nascent neuroimaging of deception. *Neurocase, 14*, 68-81.

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

Squire, L.R., van der Horst, A.S., McDuff, S.G.R., Frascino, J.C., Hopkins, R.O., & Mauldin, K.N. (2010). Role of the hippocampus in remembering the past and imagining the future. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 19044-19048.

Szpunar, K. K. (2010). Episodic future thought: An emerging concept. *Perspectives on Psychological Science, 5,* 142-162.

Szpunar, K.K., Chan, J.C.K., & McDemott, K.B. (2009). Contextual processing in episodic future thought. *Cerebral Cortex, 19,* 1539-1548.

Szpunar, K. 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,* 642-647.

Treyer, V., Buck, A., & Schnider, A. (2003). Subcortical loop activation during selection of currently relevant memories. *Journal of Cognitive Neuroscience, 15,* 610-618.

Tulving, E., and Schacter, D.L. (1990). Priming and human memory systems. *Science, 247,* 301-306.

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13,* 230–241.

Verschuere, B., Prati, V., & De Houwer, J. (2009). Cheating the lie detector: Faking in the autobiographical impliciat association test. *Psychological Science 20,* 410-413.

Von Zerssen, G.C., Mecklinger, A., Opitz, B., & von Cramon, D.Y (2001). Conscious recollection and illusory recognition: an event-related fMRI study. *European Journal of Neuroscience, 13,* 1-13.

Wagner, A. D. (2010). Can neuroscience identify lies? In M. S. Gazzaniga & J. S. Rakoff (Eds.), *A judge's guide to neuroscience: A concise introduction* (pp. 13-25). Santa Barbara: University of California, Santa Barbara.

Wagner, A.D., Paré-Blagoev, E.J., Clark, J., and Poldrack, R.A. 2001. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31,* 329-338.

Walla, P., Endl, W., Lindinger, G., Deecke, L., & Lang, W. (2000). False recognition in a verbal memory task: an event-related potential study. *Cognitive Brain Research, 9,* 41-44.

Weiler, J.A., Suchan, B., & Daum, I. (2010). When the future becomes the past: Differences in brain activity patterns for episodic memory and episodic future thinking. *Behavioural Brain Research, 212,* 196-203.

Wells, G.L. & Olson, E.A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54,* 277-295.

Welss, G.L., Small, M., Penrod, S., Malpass, R.S., Fulero, S.M., & Brimacombe, C.A.E. (1998). *Law and Human Behavior, 22,* 603-647.

Wheeler, M. A., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid recollection activates modality-specific cortex. *Proceedings of the National Academy of Sciences of the United States of America, 97,* 11125-11129.

Wiggs, C.L., and Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology, 8,* 227-233.

Wise, R. J. S., & Price, C. J. (2006). Functional imaging of language. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (2nd ed., pp. 191–228). Cambridge, MA: MIT Press.

Acknowledgments

Preparation of this chapter was supported by grants from NIMH and NIA. We thank

Clifford Robbins for assistance with preparation of the manuscript and Nobuhito Abe for helpful

comments on an earlier draft of the chapter.