# False memories and confabulation

**Marcia K. Johnson and Carol L. Raye**

**Memory distortions range from the benign (thinking you mailed a check that you only thought about mailing), to the serious (confusing what you heard after a crime with what you actually saw), to the fantastic (claiming you piloted a spaceship). We review theoretical ideas and empirical evidence about the source monitoring processes underlying both true and false memories. Neuropsychological studies show that certain forms of brain damage (such as combined frontal and medial-temporal lesions) might result in profound source confusions, called confabulations. Neuroimaging techniques provide new evidence regarding more specific links between underlying brain mechanisms and the normal cognitive processes involved in evaluating memories. One hypothesis is that the right prefrontal cortex (PFC) subserves heuristic judgments based on easily assessed qualities (such as familiarity or perceptual detail) and the left PFC (or the right and left PFC together) subserves more systematic judgments requiring more careful analysis of memorial qualities or retrieval and evaluation of additional supporting or disconfirming information. Such heuristic and systematic processes can be disrupted not only by brain damage but also, for example, by hypnosis, social demands and motivational factors, suggesting caution in the methods used by 'memory exploring' professions (therapists, police officers, lawyers, etc.) in order to avoid inducing false memories.**

There is nothing that seems to belong to us so much as our personal memories. When two people disagree about how they remember a shared autobiographical event, often, each is quite sure the other is wrong. The idea that memories for personally significant episodes in our life might be false raises a disturbing question: can we trust any of our memories? To what extent do we confabulate our lives? And what is the difference between normal distortions of memory and clinical confabulation and delusion?

### False memories

All experience is constructed in that people use their general knowledge of the world to fill in 'missing elements' (elements that are not there or that people fail to notice) during perception and then again when they remember events later[1,2]. Of course, if people know that they have filled in certain information we do not say that their memory is distorted, only that they are making inferences given incomplete information. However, when they mistake what they only thought for what actually happened[3], then the memory is distorted, false or illusory – there has been a failure of 'reality monitoring', a failure to distinguish between perceived information and internally generated information in memory[4]. Sometimes, people not only confuse the real and the imagined, or actual events and their knowledge and beliefs (such as schemas and stereotypes), but they confuse elements from various perceived events (such as television news and a fictional novel). The terms 'source confusions' ('misattributions')[5,6] include both source and reality monitoring errors, which are responsible for many memory distortions. By using simplified situations and studying source confusions in the laboratory, researchers have characterized a number of factors and processes that underlie both true and false memories. We describe these in an approach to understanding memory processes called the source monitoring framework (SMF)[5].

First, construction does not imply necessarily that any information was lost from memory; for example, perceptual representations and our constructed understanding of events might both persist in memory and be reflected in performance under different conditions[7,8]. Second, there is no single piece of information that invariably marks a memory as an accurate reflection of the past. But there are several ways that we can and do monitor (or evaluate) what we remember; some of these compare the features of a memory to a typical pattern; others rely on more deliberative processes. On average, memories from different sources differ in their phenomenal qualities and people make heuristic source attributions based on such differences. Typically, memories for experienced (external) events have information denoting time, location, spatial arrangement, emotion or sensory perceptual details such as color and shape. In contrast, memories for thoughts and imagined events typically have much less or less vivid information of these types, but often have more information about cognitive operations (such as intention

*M.K. Johnson and C.L. Raye are at the Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544-1010, USA.*

**tel: +1 609 258 4664**
**fax: +1 609 258 1113**
**e-mail: mkj@clarity.princeton.edu**

 1364-6613/98/$19.00 PII: S1364-6613(98)01152-8

## Box 1. Features from one memory affect source attributions about another

In a study by Henkel *et al.*[a], in each trial, participants saw either a word accompanied by a picture or they saw a word only and were required to imagine a corresponding picture. Imagined items were either physically similar to a perceived item on the list (for example, a lollipop and a magnifying glass), conceptually similar to a perceived item (for example, a banana and an apple) or were unrelated to perceived items (for example, a screwdriver and a hanger). Later, participants received a surprise source monitoring test: older adults (mean age, 74 years) were tested after a 15-min retention interval and younger adults (mean age, 20 years) were tested after a two day interval to equate groups on old/new recognition; another older group was tested after two days as a comparison with young subjects after a comparable retention interval. On the test, subjects saw words that accompanied previously presented or imagined pictures or new words, and responded with: 'picture', 'imagined' or 'new' to each test word. Scoring simply for old/new recognition (see Fig. A) of imagined items (regardless of whether the source attribution was correct), older adults tested at 15 min performed at the same level as younger adults tested after two days. Older adults tested at two days had significantly lower old/new recognition, but were still showing substantial memory. On source identification (Fig. B) for unrelated items, older adults tested at 15 min were just as accurate as the younger adults in identifying the origin of an imagined item, but for physically or conceptually similar items older adults were more likely to claim to have seen an imagined item. Thus, older adults were not simply more confused in general but suffered in particular from similarity among features of candidate memories. Older adults tested at two days were at chance on source identification, even though their old/new recognition was well above chance. Older adults also took neuropsychological test batteries to assess medial-temporal function and frontal function (age-related neuropathology has been found in both of these regions). Source accuracy was correlated with medial-temporal scores at both the 15-min and two-day test and with frontal scores on the two day test (see also Refs b, c). These results are consistent with the idea that medial-temporal and frontal regions play somewhat different roles in source memory, with medial-temporal regions more important for binding features into complex memories and frontal regions more important for strategic retrieval and evaluation[d,e].

**References**

**a** Henkel, L.A., Johnson, M.K. and De Leonardis, D.M. Aging and source monitoring: cognitive processes and neuropsychological correlates *J. Exp. Psychol. Gen.* (in press)

**b** Craik, F.I.M. *et al.* (1990) Relations between source amnesia and frontal lobe functioning in older adults *Psychol. Aging* 5, 148–151

**c** Glisky, E.L., Polster, M.R. and Routhieaux, B.C. (1995) Double dissociation between item and source memory *Neuropsychology* 9, 229–235

**d** Johnson, M.K., Hashtroudi, S. and Lindsay, D.S. (1993) Source monitoring *Psychol. Bull.* 114, 3–28

**e** Moscovitch, M. (1995) Confabulation, in *Memory Distortion* (Schacter, D.L., ed.), pp. 226–251, Harvard University Press

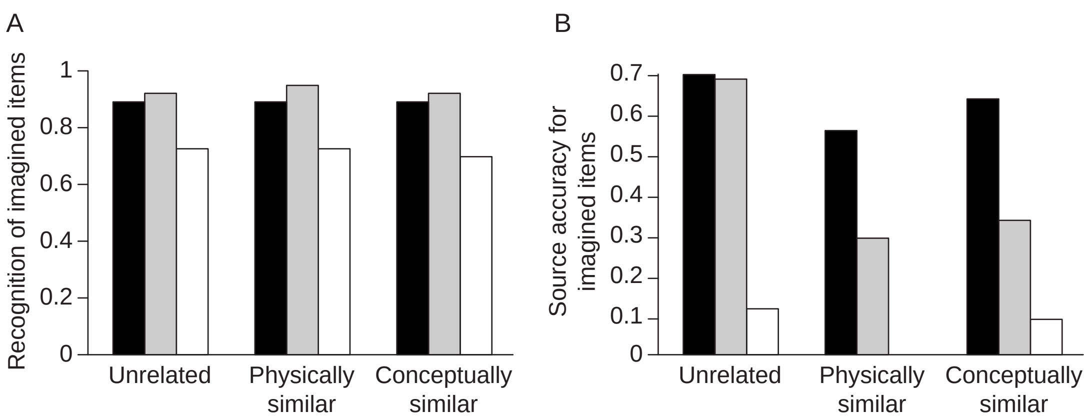


**Fig. Performance on a source memory test scored for recognition (A) or source identification (B) of previously imagined items.** Recognition of imagined items (A), was similar for young adults tested two days after the initial trials (black bars) and older adults tested after 15 min (grey bars), but a second group of older adults tested after two days (white bars) showed lower recognition rates. When measured on source identification of test items (B) young and old (15-min test) groups were equally accurate for test items that were unrelated, but older adults were less accurate for physically or conceptually related items. Older adults tested after two days performed at chance level on all test categories. (Modified from Ref. a.)

and planning, deliberate imaging, actively searching for a piece of information and drawing conclusions). Similarly, on average, the memories from a specific external source (television) will have different qualities (sound, motion and many visual details) than the memories from another source (the newspaper). Memory monitoring processes capitalize on such differences by evaluating memories (or mental experiences in general) for their match with the expected characteristics of a given source. Such attributions are correct sufficiently often to keep our memories and beliefs constrained by reality, but, importantly, are subject to error. For example, the information in the memory may be vague owing to poor encoding or retrieval processes; the comparison pattern is only probabilistic information (for example, some memories for imagined events are more perceptually detailed than some memories for perceived events and, hence, might be mistaken for perceived events); or the comparison information used (such as simple familiarity) or the decision criterion (strict or lax) might be inappropriate for the current task.

## Box 2. The potential memory-distorting effects of forced confabulation

Ackil and Zaragoza[a] showed children and college students a short excerpt from a movie depicting a boy's experiences at a summer camp. After the film, students were asked questions about the film and were told that they must give an answer to every question, even if they had to guess. For example, a false-event question was: 'What did the boy say Sullivan had stolen?' when Sullivan was not shown stealing anything, nor was there any reference to his stealing anything. A week later, students were asked whether they had seen various items in the film. The results showed that all age groups were more likely to misattribute items that they had generated (as forced responses to questions about the film) than they were to misattribute control items (previous guesses by other students) to the film (see Fig.). This memory-distorting effect of forced confabulation was greater for children than for the college students. This effect was obtained even though participants had been reluctant to guess initially and even though they were told before the test that the experimenter from the previous week had made a mistake and asked them about some things that did not happen in the film. For example, here is the transcript from the initial questioning of a third-grader:

Experimenter: 'What did the boy say Sullivan had stolen?'
Participant: 'Ahh, I forget what that was.'
Experimenter: 'Oh, can you just take a guess then?'
Participant: 'Mmm, no, I don't think so.'
Experimenter: 'Well, what do you think would make him really mad if Sullivan had stolen it?'
Participant: 'Ahh, maybe like a radio or something?'
Experimenter: 'OK.'

A week later, this student claimed to have seen the confabulated incident regarding the stolen radio in the film.

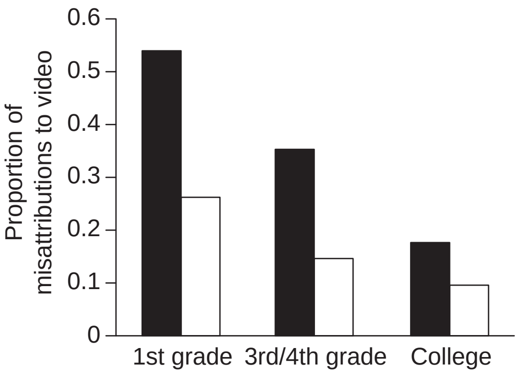


**Fig. Memory distortion resulting from forced confabulation.** A week after watching a short film, first and third/fourth grade school children and college students both misattribute a higher proportion of items generated by themselves as forced responses (black bars) than control items generated by other students (white bars). The effect is greater, however, for the children than for college students. (Modified from Ref. a.)

**Reference**

**a** Ackil, J.K. and Zaragoza, M.S. The memorial consequences of forced confabulation: age differences in susceptibility to false memories *Dev. Psychol.* (in press)

Other, more deliberative (systematic) decisions can be based on memory features or additional information, knowledge or beliefs that refute or support a particular attribution. For example, a vague memory could be judged as real if other supporting information can be retrieved (for instance, you remember a later conversation about the event). Noting some inconsistency or implausibility within a memory or between a memory and other memories or knowledge indicates some inaccuracy (for example, features of two memories might be conflated). Supporting information or consistency or plausibility provide reality constraints that often are useful, but that are not infallible. Errors can still occur; for example, people are more likely to attribute the source of memories of unfriendly than friendly behaviors to a skinhead (J.W. Sherman and G.R. Bessenoff, pers. commun.).

Similarity in features is a key factor in whether memories will be confused. For example, imagining in another person's voice increases the chances that you will confuse what you imagined they said with what they actually said[9]. Perceptual or semantic features from a memory of a perceived event that are activated along with a memory of a similar imagined event can make the imagined event seem perceived (Box 1). In fact, false memories that an item (such as a needle) occurred previously are likely if many items that were presented are semantically related to it (for example, haystack, sharp or thread)[10] or have similar perceptual features (for instance, seeing pictures of many different shoes induces later false recognition of a new type of shoe)[11]. When events are highly similar, it is critical that encoding and consolidation processes incorporate (or bind) distinctive features into an episodic representation.

The information we fill in without being aware of it (the associations and inferences that we make) is perhaps the information most likely to be remembered falsely. Yet, even when we know at the time that we are engaged in construction, we may later mistake what we generated for what happened[12] (Box 2). Over time, the cognitive operations features that told us initially that we had generated information degrade and become less distinctive, or we may ignore or give such information less weight, basing source attributions instead on other information (such as familiarity, perceptual detail, supporting memories or plausibility).

Information that simply tells us something was previously experienced often revives more quickly than information that allows us to identify its source[13,14]. Often, remembering requires sustained attention to unfolding associations and requires a reasonable strategy for self-generation of likely cues[15–17]. Furthermore, as in perception, in making memory attributions people attend selectively to some information and ignore other information. For example, if people see a picture and later read a description of the scene with some misinformation (for instance, that there was a hat-rack in the scene), later they are likely to recognize falsely that there was a hat-rack in the picture. However, if instead of being asked whether they saw the hat-rack in the picture (yes or no), they are asked to indicate whether they saw, read about, both saw and read

## Box 3. Is neural activity different for true and false memories?

In one type of source monitoring study, participants hear lists of associates to non-presented 'lure' words (for example, thread, pin, eye, sewing, sharp, pin, prick, thimble, haystack and thorn, all of which are associates of needle) and are very likely to recognize falsely the non-presented lures on a subsequent recognition test[a]. Does this mean that neural activity is the same for true and false memories? The answer depends on how people are tested. Johnson *et al.*[b] recorded event-related potentials (ERPs) from the scalp of participants during the recognition test. In one condition, test items were presented in a 'blocked' fashion (several old items, followed by several lures, followed by several new items, etc.) and in another condition, old items, lures and new items were randomly intermixed. ERPs (here illustrated for frontal sites) for correct responses to old items and false recognition of lures were different in the blocked condition but strikingly similar in the random condition (see Fig.). Johnson *et al.* suggested that in the random condition, participants respond on the basis of semantic familiarity; hence, ERPs to lures looked much like ERPs to old items. In the blocked condition, familiarity would not be as easy to use because successive items within blocks would have about the same familiarity; hence, participants consider each item more carefully, attempting to find a basis for discriminating among similar items within a block. This closer evaluation is reflected in the ERPs, which evidently reflect differences in qualitative characteristics of true and false memories[c,d]. In short, whether it appears that 'true' and 'false' memories are the same or different depends not only on the nature of the encoded information, but also on what information subjects are induced to consider at test. Note that even under conditions where true and false memories do show differences, they are only different on average (that is, the distributions of features in true and false memories overlap). Thus, brain activity is unlikely to provide a sure-fire way of identifying false memories.

**References**

**a** Roediger, H.L. and McDermott, K.B. (1995) Creating false memories: remembering words not presented in lists *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814

**b** Johnson, M.K. *et al.* (1997) The similarity of brain activity associated with true and false recognition memory depends on test format *Psychol. Sci.* 8, 250–257

**c** Mather, M., Henkel, L.A. and Johnson, M.K. (1997) Evaluating characteristics of false memories: remember/know judgments and memory characteristics questionnaire compared *Mem. Cognit.* 25, 826–837

**d** Norman, K.A. and Schacter, D.L. (1997) False recognition in younger and older adults: exploring characteristics of illusory memories *Mem. Cognit.* 25, 838–848

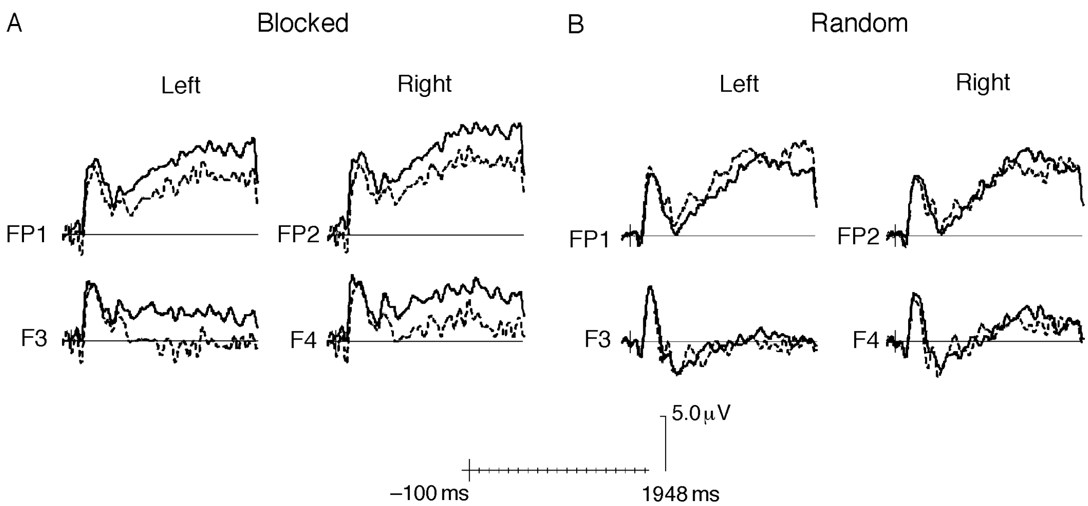


**Fig. Event-related potentials (ERPs) recorded during a word recognition test.** Responses correlated with recognition of previously heard words (unbroken lines) could be differentiated from those corresponding to false recognition of 'lure' words (broken lines) when the test words were presented in blocks **(A)**, but not when test items were randomly intermixed **(B)**. (Modified from Ref. b.)

about, or neither saw nor read about the hat-rack, they are less likely to claim falsely that they have seen it. With the first question, it appears that subjects do not use all the information that they have; they might say that they have seen it based on its familiarity or plausibility, or perhaps based on a relatively low level of (imagined) perceptual detail in the hat-rack memory. The second type of question induces people to retrieve and/or examine more carefully the information available, particularly features that would allow them to distinguish between possible sources[18] (Box 3).

### Confabulation

As can be seen from the above discussion, distortions of memory are a byproduct of normal memory processes. However, sometimes, as a consequence of brain damage, memory distortions occur that are clearly beyond this normal range, either because of their increased incidence, or their bizarreness. Such confabulations are clinically significant false statements that patients make without intention to deceive. For example, Benson *et al.* reported a patient who gave detailed descriptions of conversations with physicians that she had never met and trips she had made out of the hospital that had not occurred[19]. Stuss *et al.* reported a patient who fabricated a story of a drowning accident in which he rescued one of his children and another patient who fabricated stories about how members of his family had been killed before his eyes[20]. Damasio *et al.* described a patient who claimed to have been a pirate in charge of a spaceship[21]. Patients confabulate about both autobiographical and non-personal, historical events[22], although the relative incidence of personal ('episodic') and non-personal ('semantic') confabulations appears to vary with the patient and the method of testing[23,24]. Confabulations vary in plausibility from relatively mild – for example, filling in of gaps, loose paraphrasing and temporal displacements of actual events – to more severe, highly implausible and bizarre accounts[25–27]. Some confabulations have qualities similar to those of real memories[28,29].

In theory, confabulations could be produced by disruption of any of the processes outlined in the SMF that are critical for normal memory. That is, confabulations might reflect poor memory as a result of: (1) inadequate feature binding, (2) disrupted reactivation and consolidation processes, (3) failure to engage evaluation processes or to use situationally appropriate feature weights and criteria, (4) poor self-initiated cuing and retrieval of specific related supporting/disconfirming information, or (5) failure to access or use general knowledge about the world or the self to constrain source attributions in ways that preclude fantastic or bizarre memories[30]. Deficits in such mechanisms might be compounded by motivational and personality factors or individual differences in vividness of imagery[23,29,31]. Most theories of confabulation incorporate one or more of these factors[17,20,30,32,33]. Different combinations of deficits might account for the variety of confabulation syndromes, and the more processes that are disrupted, the greater should be the chance for more severe, fantastical confabulations[22,23,30].

## Brain mechanisms of source memory

### *Evidence from patients*

Based on patient data, two types of constraints on source monitoring errors suggested by the SMF can be associated with general brain regions: (1) The first type of constraint comes from memory representations in which features are well bound and thus well differentiated from other representations with which they might be confused[34,35]. Damage in medial-temporal regions (especially the hippocampal formation), diencephalic regions or the basal forebrain can result in amnesia for new events and information experienced after the damage (anterograde amnesia). These structures appear to be part of neural circuits necessary for binding, reactivating and consolidating the various features of experience into complex, distinguishable, episodic memories[36,37]. (2) The second type of constraint comes from heuristic and systematic evaluation and retrieval processes that attribute activated information to sources[4,17]. Damage in frontal regions is associated with deficits in self-initiated (or executive) processes such as generating cues for retrieval, switching sets, monitoring the appropriateness of responses (including adopting evaluative standards relevant to the task and setting criteria), inhibiting inappropriate responses and temporal and other source judgments[35,38,39].

However, typically, neither memory deficits (not even profound amnesia) nor executive deficits alone produce confabulation. Many amnesic patients never show clinically significant confabulation, and many who confabulate for a period post-trauma show a marked decrease in confabulation without any appreciable improvement in memory[19,40]. In one particularly striking case, O'Connor *et al.* described a patient who had severe amnesia for many years with no history of confabulation, who then began to confabulate after a closed head injury (Box 4). With respect to executive deficits, scores on various neuropsychological tests assessing frontal function are not always correlated among themselves; it is possible to have poor performance on one or more frontal tests without clear evidence of frontal damage from a scan, and patients with profound frontal damage, indicated by both neuropsychological testing and brain scans, do not necessarily confabulate[23].

The most florid confabulation is associated with frontal damage combined with memory deficits[20,41]. Important evidence comes from observations of confabulation in patients who have suffered aneurysm of the anterior communicating artery (ACoA)[22,40,42–44]. The neuropathological consequences of ACoA are varied, but the basal forebrain is supplied by the anterior communicating artery; thus, damage to basal forebrain structures, along with damage to ventromedial or orbitomedial frontal regions (often bilateral) is common. ACoA patients exhibit both amnesia and executive deficits[45]. Confabulation is also sometimes seen in alcoholic Korsakoff patients[19], often in a relatively mild form[27]. Korsakoff patients show damage (often bilateral) to the diencephalon and frontal lobes and exhibit amnesia and executive deficits[45]. Benson *et al.* reported an interesting SPECT (single-photon emission computerized tomography) study of a Korsakoff patient who showed hypoperfusion in the medial orbital frontal area and cingulate [but not the dorsolateral prefrontal cortex (PFC)] during the period of confabulation, which improved as confabulation decreased[19]. (See Box 5 for issues that arise in collecting confabulation data.)

Because aging increases the chances of neuropathology in both frontal and medial-temporal regions, older adults provide another (and larger) population for the study of neural mechanisms of source monitoring[46,47]. Reports show correlations between source accuracy and scores on frontal tests in older adults, and more recent evidence shows correlations with scores on medial-temporal tests as well (Box 1).

### *Evidence from neuroimaging*

While studies of neurologically impaired patients remain critical, an increasingly important source of evidence regarding brain mechanisms of veridical and false memory comes from electrophysiological and functional imaging [positron emission tomography (PET) and functional magnetic resonance imaging (fMRI)] studies. The earliest findings focused attention on activity of the right prefrontal cortex during the test phase of episodic memory tasks[48–50]. Several hypotheses about this right PFC activity have been suggested: that it reflects a retrieval 'mode' engaged during episodic tests[51], retrieval effort[52] or evaluation[50]. Recently, Johnson *et al.* reported electrophysiological evidence supporting the idea that the right frontal activity reflects evaluation of activated information[53] (also, see Refs 54–56). Consistent with the neuroimaging data, Johnson *et al.* found more positive event-related potential (ERP) amplitudes at right than left electrodes for anterior but not posterior electrode sites. Four features of their data led them to suggest that the right frontal activity seen in neuroimaging

## Box 4. Delusional misidentification: Capgras syndrome

Variations in brain areas that are damaged produce different types of confabulatory/delusional syndromes. For example, O'Connor *et al.* described a well-adjusted patient (EB) who had amnesia for over ten years and then suffered a closed head injury from a fall[a]. A computed tomography scan following her injury showed a right temporoparietal hematoma, and neuropsychological testing showed much poorer performance on frontal tests than before her accident. EB began to confabulate, and developed the delusion that her husband was her father (who had been dead for 20 years). This kind of misidentification of person (Capgras) syndrome has been associated with right hemisphere lesions that disrupt visuospatial processing, together with frontal dysfunction and/or memory deficits. Ellis *et al.* suggest that two types of impairment may be necessary to produce the Capgras delusion, a cognitive/affective error and a misattribution error[b]. They propose that disruption of pathways that mediate emotional responses to personally significant visual stimuli, primarily faces, underlies the cognitive/affective error and that impairment of judgment processes might also be necessary[c]. Although EB's visuospatial test scores did not change after her fall, her recognition of famous faces declined dramatically. EB was aware that her father had died and could admit that her father and husband did not look alike, but she was not able to use consistency or plausibility information to attribute source correctly, or she gave greater weight to her feelings. Perhaps EB's delusion was an attempt to make sense of an unusual affective response to her husband. As a child, EB had been sexually molested by her father, and she was extremely anxious around her husband when she was delusional. Thus, brain damage might set the conditions for confabulation, but the patient's history may determine the content. Fortunately, EB's frontal scores improved eventually and her delusions subsided during this same time period. The combination of frontal damage and memory deficits in confabulation and similar delusional syndromes is also seen in patients with Alzheimer's disease (AD) according to a recent postron emission tomography (PET) study[d]. Relative to those without, those AD patients with delusional misidentification syndromes showed hypometabolism in orbitofrontal, cingulate, and left medial-temporal areas (areas also damaged in other confabulating patients – see text), and hypermetabolism in sensory association cortices. Clearly, a task at hand is to specify more exactly how different areas of the brain interact with the prefrontal cortex to produce various types of misattributions of mental experiences.

**References**

**a** O'Connor, M. *et al.* (1996) A neuropsychological analysis of Capgras syndrome *Neuropsychiatry Neuropsychol. Behav. Neurol.* 9, 265–271

**b** Ellis, H.D. *et al.* (1997) Response from Ellis, Young, Quale and de Pauw *Trends Cognit. Sci.* 1, 158

**c** Ellis, H.D. *et al.* (1997) Reduced autonomic responses to faces in Capgras delusion *Proc. R. Soc. London Ser. B* 264, 1085–1092

**d** Mentis, M.J. *et al.* (1995) Abnormal brain glucose metabolism in the delusional misidentification syndromes: a positron emission tomography study in Alzheimer disease *Biol. Psychiatry* 38, 438–449

reflects evaluation: (1) the right–left asymmetry was not greater for source identification than for old/new recognition, even though source identification requires more episodic detail; (2) the right–left asymmetry emerged over time during individual trials – 700 ms and more after the test item was presented; (3) an earlier, bilateral ERP component occurring ~400 ms after the test item was correlated with the type of features (visual or non-visual) accessed; and (4) an additional component, observed for left as well as right frontal electrode sites, differentiated source identification from old/new recognition. Johnson *et al.* suggested that the right PFC is involved in heuristic evaluation processes necessary for both recognition and source identification, and that the left and right PFC are involved jointly in more effortful systematic processes, including reflectively guided retrieval, which is more necessary for source attributions[57] (Burgess and Shallice propose a similar model[17]). Consistent with this, unilateral damage to either the right[58] or left[59] PFC can produce deficits in source monitoring.

Although strong claims about the localization of neural activity cannot be made from these ERP data, in general, results of PET and fMRI studies are consistent with the heuristic/systematic hypothesis based on the SMF. Significant right but not left PFC activity tends to be found for relatively simple episodic test tasks, such as forced-choice recognition[60] or easy cued recall (for example, poet–Browning)[50]. Typically, more complex episodic test tasks, such as those with fragment cues ( E_G_E for EAGLE)[61] or free recall[62], produce left as well as right PFC activity. Furthermore, studies that directly compare difficult recognition with relatively easy recognition[63], or compare difficult cued recall with relatively easy cued recall[64], tend to find greater activity in the left (as well as the right) PFC for the more difficult conditions.

Nevertheless, one puzzle is that the brain regions implicated in imaging studies are not, generally speaking, the brain regions implicated most often in studies of confabulating patients. In the imaging studies, the regions (right or right and left) showing significant activation tend to be dorsolateral PFC areas (most often Brodmann's areas 10 and 46, but also areas 9, 44, 45 and 47), not the orbitomedial or ventromedial areas implicated in confabulation. This could be because orbitomedial and ventromedial areas are not as likely to be scanned, or because they are involved in baseline as well as experimental tasks and do not show up in subtractions. Newer, trial-based imaging analysis techniques[65], where activity can be compared with intertrial rest periods, should help in detecting regions that are active when individuals are engaged in goal-directed tasks.

In any event, an important next step is to specify the relative contributions of medial and lateral PFC regions[62] and the nature of their interactions. One possibility is that damage to medial areas disrupts pathways critical for dorsolateral PFC functioning. Another is that one or more of the medial regions implicated in confabulation (orbitomedial/ventromedial or anterior cingulate) subserve monitoring processes, or they modulate or coordinate activity in lateral PFC regions. Another is that medial damage is especially

## Box 5. Collecting data on confabulation

Collecting systematic data on confabulating patients is difficult. First, there is the difficulty of quantifying and characterizing confabulation. It would be helpful to develop standard methods for eliciting and describing patients' confabulations[a,b]. Also, there is considerable variation in: (1) how long after damage occurs that patients are evaluated, (2) the nature of the neuropsychological tests given, and (3) the information about how long confabulation persists in a given patient. Standardization would facilitate comparisons across studies of patients with different neuropathology[c]. It would be useful to test confabulating and non-confabulating patients with similar brain damage and/or neuropsychological profiles on a range of experimental tasks directed at more specific aspects of source memory, such as retrieval requiring self-generation of cues and identification of various attributes of source (temporal, spatial, person, etc.), collecting both patients' and experimenters' ratings of qualitative characteristics of memories[d]. Finally, of course, specific, detailed information about damaged brain regions is critical, ideally expressed in a standard format[e].

**References**

a Dalla Barba, G. (1993) Confabulation: knowledge and recollective experience *Cognit. Neuropsychol.* 10, 1–20
b Moscovitch, M. and Melo, B. (1997) Strategic retrieval and the frontal lobes: evidence from confabulation and amnesia *Neuropsychologia* 35, 1017–1034
c Squire, L.R. and Shimamura, A.P. (1986) Characterizing amnesic patients for neurobehavioral study *Behav. Neurosci.* 100, 866–877
d Johnson, M.K., O'Connor, M. and Cantor, J. (1997) Confabulation, memory deficits, and frontal dysfunction *Brain Cognit.* 34, 189–206
e Swick, D.K. and Knight, R.T. (1996) Is prefrontal cortex involved in cued recall? A neuropsychological test of PET findings *Neuropsychologia* 34, 1019–1028

likely to disrupt reflective functions bilaterally and certain types of retrieval and monitoring (such as systematic) might depend on both hemispheres or an interaction between them. In this regard, the occasional observations in neuroimaging studies of episodic memory of activation in the cingulate are particularly interesting. Finally, confabulation is most likely to occur with combined binding deficits (amnesia) and damage to frontal areas subserving source monitoring processes; compared with orbitomedial/ventromedial damage, direct dorsolateral PFC damage is less likely to co-occur in combination with damage that produces amnesia.

In addition, it should be noted that although memory and executive systems are usually considered separately, they work together during both encoding and remembering. For example, binding may be a function of the medial-temporal system but opportunities for binding are undoubtedly controlled by frontally maintained goals or agendas as well as encoding strategies. Another reason to consider the interactions between memory and executive systems comes from considering the possible role of neurotransmitters, especially dopamine and acetylcholine, in source memory. Antipsychotic drugs that reduce positive symptoms (hallucinations and delusions) in schizophrenic patients modulate dopamine. D'Esposito and Alexander suggest that dopamine is important for modulating short-term memory and that acetylcholine might be critical for long-term episodic memory[66]. The concept of short-term memory includes the kind of executive or reflective PFC functions critical for encoding, retrieving and evaluating memories[57]. Damage to the basal forebrain and adjacent areas is likely to disrupt the functions of neurotransmitters, including acetylcholine and dopamine, which originate in the basal forebrain and brainstem and project to PFC and medial-temporal areas[67]. Thus, fluctuations in dopamine and acetylcholine, for example, as a result of brain damage, might be expected to affect source monitoring.

### Applications

The SMF provides a theoretical context for understanding the potential impact of interviewing techniques in forensic contexts such as eyewitness testimony[68,69], including children's testimony in potential cases of sexual abuse[70]. It also accounts for the potential to induce false memories and beliefs during therapy, for example, in attempting to recover repressed memories[71] or communicate with multiple personalities[72]. For example, recent surveys indicate that a substantial number of therapists use techniques (such as guided imagery, hypnosis or interpretation of dreams) to help clients recover suspected memories of childhood abuse – techniques that have the potential for inducing false memories[73]. For instance, hypnosis enhances confidence without enhancing memory, perhaps by shifting the weights assigned to different qualities of memory by evaluation processes[74]. More generally, basic source monitoring principles have been shown to hold in a number of complex, naturalistic situations. If people are exposed to a bystander before a 'lineup', the familiarity of the bystander might be misattributed to familiarity from the witnessed crime[75]. If possible items or events are suggested during questioning, they might later be remembered falsely as part of the target event[76], and repeated questioning about something that did not occur is likely to increase false memories[77]. Forcing people to knowingly confabulate answers to questions can lead later to false memories for the confabulated information, especially in children (Box 2). By leading people to imagine events or to generate supporting memories they can be induced to 'remember' elaborate autobiographical episodes that did not occur[78,79].

In addition to the cognitive, motivational, personality and social factors that might induce memory distortion in situations such as forensic interviews and therapy, possible neurobiological factors are of interest too. For example, children's susceptibility to source misattributions might be related to the degree to which their frontal regions have developed[80], and the susceptibility of both children and adults might be affected by conditions that affect the levels of neurotransmitters that modulate frontal functions.

### Conclusion

The question posed at the start of this article was: can we trust our memories? The answer is that, in fact, we have to and

## Outstanding questions

- What neural systems support the many different cognitive processes necessary for normal source monitoring? Can we design cognitive tasks for use in combination with neuroimaging techniques to clarify functions of the different frontal regions (ventromedial, orbitomedial, dorsolateral, ventrolateral)? How do these interact with distributed sensory-association areas and medial-temporal (and other memory) structures in encoding and consolidating multiple features of an event? To what extent are these same systems and other systems involved in retrieval and evaluation of memories?
- What are the important differences between patients with similar etiology [for example, patients with aneurism of the anterior communicating artery (ACoA)] who do and do not confabulate in terms of their performance on cognitive tasks that assess different factors that influence source monitoring? Which standard neuropsychological test battery components correlate with these factors and is it possible to develop a battery that discriminates among types of confabulations and duration of confabulation?
- What is (are) the primary effect(s) of ACoA basal forebrain damage that predict confabulation? What structures and systems are involved? What is the best way to organize a combined neuroimaging, pharmacological and cognitive/behavioral approach to understanding confabulation? What is the best way to facilitate sharing patient information (for instance, the results of brain scans, standard neuropsychological tests and experimental measures)?
- How do the cognitive and neural mechanisms of memory interact with emotion?
- How do children develop source monitoring criteria as part of their understanding of their own (and others') mental experience? What could we learn from systematically comparing their cognitive, social and neural development with respect to source monitoring?

we do – but, importantly, we trust our source monitoring processes to indicate not only when memories probably correspond to reality but also when they might not do so. Various processes operate to constrain the amount of distortion that arises from our imperfect memory system and to signal us when we should be cautious about the truthfulness of a memory. The feeling of remembering is important to our well-being, but so is the feeling of not remembering that accompanies vague, inconsistent, or implausible recollections. Accurate memory is knowing when we do not remember as well as knowing when we do. The subtle balance of the encoding, consolidation, reactivation, retrieval and evaluation processes that underlie the 'meta-memory' function of source monitoring develops throughout childhood, tends to weaken in old age, and can be disrupted at any age by distraction, stress, drugs, hypnosis and social or motivational pressures. A profound disorganization of memories and beliefs can occur when memory monitoring processes are disrupted as a consequence of frontal brain damage, especially in combination with memory deficits from damage to medial-temporal, diencephalic or basal forebrain areas. The SMF includes the working hypothesis that the right frontal PFC is involved in relatively easy, heuristic evaluation processes and that the right and left PFC cooperate and coordinate in more reflectively demanding (systematic) retrieval and evaluation processes. Further constraints are placed on source monitoring by the nature of the information available for evaluation. Hence, deficits in perceptual or memorial processes (for example, feature binding) will increase the chances of confabulation and delusions. Electrophysiological and brain imaging studies have the exciting potential to help us link more particular brain regions within the PFC (for example, the ventromedial, orbitomedial and dorsolateral areas) and more particular structures within memory circuits (for example, the hippocampus, dorsomedial nucleus of the thalamus and septal area) to specific component processes of remembering and map out the relations among them and other brain regions in which the records of our experiences (both real and imagined) are represented.

**Acknowledgement**
This work was supported by NIA grant AG09253.

**References**

1 Bartlett, F.C. (1932) *Remembering: a Study in Experimental and Social Psychology*, Cambridge University Press
2 Bransford, J.D. and Johnson, M.K. (1973) Considerations of some problems of comprehension, in *Visual Information Processing* (Chase, W., ed.), pp. 383–438, Academic Press
3 Johnson, M.K., Bransford, J.D. and Solomon, S.K. (1973) Memory for tacit implications of sentences *J. Exp. Psychol.* 98, 203–205
4 Johnson, M.K. and Raye, C.L. (1981) Reality monitoring *Psychol. Rev.* 88, 67–85
5 Johnson, M.K., Hashtroudi, S. and Lindsay, D.S. (1993) Source monitoring *Psychol. Bull.* 114, 3–28
6 Zaragoza, M.S. and Lane, S.M. (1994) Source misattributions and the suggestibility of eyewitness memory *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 934–945
7 Johnson, M.K. (1981) A multiple-entry, modular memory system, in *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 17) (Bower, G.H., ed.), pp. 81–123, Erlbaum
8 Reyna, V.F. and Brainerd, C.J. (1995) Fuzzy-trace theory: an interim synthesis *Learning and Individual Differences* 7, 1–75
9 Johnson, M.K., Foley, M.A. and Leach, K. (1988) The consequences for memory of imagining in another person's voice *Mem. Cognit.* 16, 337–342
10 Roediger, H.L. and McDermott, K.B. (1995) Creating false memories: remembering words not presented in lists *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814
11 Koutstaal, W. and Schacter, D.L. (1997) Gist-based false recognition of pictures in older and younger adults *J. Mem. Lang.* 37, 555–583
12 Durso, F.T. and Johnson, M.K. (1980) The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words *J. Verbal Learn. Verbal Behav.* 19, 416–429
13 Johnson, M.K., Kounios, J. and Reeder, J.A. (1994) Time-course studies of reality monitoring and recognition *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 1409–1419
14 Gronlund, S.D., Edwards, M.B. and Ohrt, D.D. (1997) Comparison of the retrieval of item versus spatial position information *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 1261–1274
15 Reiser, B.J. (1986) The encoding and retrieval of memories of real-world experiences, in *Knowledge Structures* (Galambo, J.A., Abelson, R.P. and Black, J.B., eds), pp. 71–99, Erlbaum
16 Conway, M.A. (1992) A structural model of autobiographical memory, in *Theoretical Perspectives on Autobiographical Memory* (Conway, M.A. *et al.*, eds), pp. 167–193, Kluwer Academic Publishers
17 Burgess, P.W. and Shallice, T. (1996) Confabulation and the control of recollection *Memory* 4, 359–411
18 Lindsay, D.S. and Johnson, M.K. (1989) The eyewitness suggestibility effect and memory for source *Mem. Cognit.* 17, 349–358
19 Benson, D.F. *et al.* (1996) Neural basis of confabulation *Neurology* 46, 1239–1243
20 Stuss, D.T. *et al.* (1978) An extraordinary form of confabulation *Neurology* 28, 1166–1172
21 Damasio, A.R. *et al.* (1985) Amnesia following basal forebrain lesions *Arch. Neurol.* 42, 263–271
22 Moscovitch, M. and Melo, B. (1997) Strategic retrieval and the frontal lobes: evidence from confabulation and amnesia *Neuropsychologia* 35, 1017–1034
23 Kopelman, M.D., Ng, N. and Van Den Brouke, O. (1997) Confabulation

extending across episodic, personal, and general semantic memory *Cognit. Neuropsychol.* 14, 683–712

24 Dalla Barba, G. (1993) Confabulation: knowledge and recollective experience *Cognit. Neuropsychol.* 10, 1–20

25 Talland, G.A. (1961) Confabulation in the Wernicke–Korsakoff syndrome *J. Nerv. Ment. Dis.* 132, 361–381

26 Berlyne, N. (1972) Confabulation *Br. J. Psychiatry* 120, 31–39

27 Kopelman, M.D. (1987) Two types of confabulation *J. Neurol. Neurosurg. Psychiatry* 50, 1482–1487

28 Dalla Barba, G. (1995) Consciousness and confabulation: remembering 'another' past, in *Broken Memories: Case Studies in Memory Impairment* (Campbell, R. and Conway, M.A., eds), pp. 101–114, Blackwell

29 Johnson, M.K., O'Connor, M. and Cantor, J. (1997) Confabulation, memory deficits, and frontal dysfunction *Brain Cognit.* 34, 189–206

30 Johnson, M.K. (1991) Reality monitoring: evidence from confabulation in organic brain disease patients, in *Awareness of Deficit After Brain Injury: Clinical and Theoretical Issues* (Prigatano, G.P. and Schacter, D.L., eds), pp. 176–197, Oxford University Press

31 Weinstein, E.A. (1996) Symbolic aspects of confabulation following brain injury: influence of premorbid personality *Bull. Menninger Clin.* 60, 331–350

32 Moscovitch, M. (1995) Confabulation, in *Memory Distortion* (Schacter, D.L., ed.), pp. 226–251, Harvard University Press

33 Baddeley, A. and Wilson, B. (1986) Amnesia, autobiographical memory, and confabulation, in *Autobiographical Memory* (Rubin, D., ed.), pp. 225–252, Cambridge University Press

34 Johnson, M.K. (1992) MEM: mechanisms of recollection *J. Cogn. Neurosci.* 4, 268–280

35 Schacter, D.L., Norman, K.A. and Koutstaal, W. (1998) The cognitive neuroscience of constructive memory *Annu. Rev. Psychol.* 49, 289–318

36 Squire, L.R. and Knowlton, B.J. (1995) Memory, hippocampus, and brain systems, in *The Cognitive Neurosciences* (Gazzaniga, M.S., ed.), pp. 825–837, MIT Press

37 Kolb, B. and Whishaw, I.Q. (1990) *Fundamentals of Human Neuropsychology*, W.H. Freeman

38 Stuss, D.T. and Benson, D.F. (1986) *The Frontal Lobes*, Raven Press

39 Shimamura, A.P. (1995) Memory and frontal lobe function, in *The Cognitive Neurosciences* (Gazzaniga, M.S., ed.), pp. 803–813, MIT Press

40 DeLuca, J. and Cicerone, K.D. (1991) Confabulation following aneurysm of the anterior communicating artery *Cortex* 27, 417–423

41 Kapur, N. and Coughlan, A.K. (1980) Confabulation and frontal lobe dysfunction *J. Neurol. Neurosurg. Psychiatry* 43, 461–463

42 Fischer, R.S. *et al.* (1995) Neuropsychological and neuroanatomical correlates of confabulation *J. Clin. Exp. Neuropsychol.* 17, 20–28

43 DeLuca, J. and Diamond, B.J. (1995) Aneurysm of the anterior communicating artery: a review of neuroanatomical and neuropsychological sequelae *J. Clin. Exp. Neuropsychol.* 17, 100–121

44 Malloy, P. *et al.* (1993) The orbitomedial frontal syndrome *Arch. Clin. Neuropsychol.* 8, 185–201

45 Parkin, A.J. and Leng, N.R.C. (1993) *Neuropsychology of the Amnesic Syndrome*, Erlbaum

46 Henkel, L.A., Johnson, M.K. and De Leonardis, D.M. Aging and source monitoring: cognitive processes and neuropsychological correlates *J. Exp. Psychol. Gen.* (in press)

47 Schacter, D.L., Koutstaal, W. and Norman, K.A. (1997) False memories and aging *Trends Cognit. Sci.* 1, 229–236

48 Squire, L.R. *et al.* (1992) Activation of the hippocampus in normal humans: a functional anatomical study of memory *Proc. Natl. Acad. Sci. U. S. A.* 89, 1837–1841

49 Tulving, E. *et al.* (1994) Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings *Proc. Natl. Acad. Sci. U. S. A.* 91, 2016–2020

50 Shallice, T. *et al.* (1994) Brain regions associated with acquisition and retrieval of verbal episodic memory *Nature* 368, 633–635

51 Nyberg, L., Cabeza, R. and Tulving, E. (1996) PET studies of encoding and retrieval: the HERA model *Psychonomic Bull. Rev.* 3, 135–148

52 Schacter, D.L. *et al.* (1996) Conscious recollection and the human hippocampal formation: evidence from positron emission tomography *Proc. Natl. Acad. Sci. U. S. A.* 93, 321–325

53 Johnson, M.K., Kounios, J. and Nolde, S.F. (1996) Electrophysiological brain activity and memory source monitoring *NeuroReport* 7, 2929–2932

54 Rugg, M.D. *et al.* (1996) Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval *Brain* 119, 2073–2083

55 Wilding, E.L. and Rugg, M.D. (1996) An event-related potential study of recognition memory with and without retrieval of source *Brain* 119, 889–905

56 Schacter, D.L. *et al.* (1997) Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study *NeuroImage* 6, 259–269

57 Johnson, M.K. (1997) Identifying the origin of mental experience, in *The Mythomanias: the Nature of Deception and Self-deception* (Myslobodsky, M.S., ed.), pp. 133–180, Erlbaum

58 Curran, T. *et al.* (1997) False recognition after a right frontal lobe infarction: memory for general and specific information *Neuropsychologia* 35, 1035–1049

59 Parkin, A.J. *et al.* (1996) Pathological false alarm rates following damage to the left frontal cortex *Brain Cognit.* 32, 14–27

60 Haxby, J.V. *et al.* (1996) Face encoding and recognition in the human brain *Proc. Natl. Acad. Sci. U. S. A.* 93, 922–927

61 Blaxton, T.A. *et al.* (1996) Functional mapping of human memory using PET: comparisons of conceptual and perceptual tasks *Can. J. Exp. Psychol.* 50, 42–56

62 Petrides, M., Alivisatos, B. and Evans, A.C. (1995) Functional activation of the human ventrolateral frontal cortex during mnemonic retrieval of verbal information *Proc. Natl. Acad. Sci. U. S. A.* 92, 5803–5807

63 Buckner, R.L. *et al.* Functional-anatomic study of episodic retrieval using fMRI: retrieval effort versus retrieval success *NeuroImage* (in press)

64 Fletcher, P.C. *et al.* (1996) Brain activity during memory retrieval: the influence of imagery and semantic cueing *Brain* 119, 1587–1596

65 Zarahn, E., Aguirre, G.K. and D'Esposito, M. (1997) A trial based experimental design for functional MRI *NeuroImage* 6, 122–138

66 D'Esposito, M. and Alexander, M.P. (1995) The clinical profiles, recovery and rehabilitation of memory disorders *Neurorehabilitation* 5, 141–159

67 Nolte, J. (1993) *The Human Brain: an Introduction to its Functional Anatomy*, Mosby-Year Book

68 Zaragoza, M.S. *et al.* (1997) Confusing real and suggested memories: source monitoring and eyewitness suggestibility, in *Memory for Everyday and Emotional Events* (Stein, N. *et al.*, eds), pp. 401–425, Erlbaum

69 Belli, R. and Loftus, E. (1994) Recovered memories of childhood abuse: a source monitoring perspective, in *Dissociation: Clinical and Theoretical Perspectives* (Lynn, S.J. and Rhue, J.W., eds), pp. 415–433, Guilford Press

70 Ceci, S.J. and Bruck, M. (1993) Suggestibility of the child witness: a historical review and synthesis *Psychol. Bull.* 113, 403–439

71 Lindsay, D.S. and Read, J.D. (1994) Psychotherapy and memories of childhood sexual abuse *Appl. Cognit. Psychol.* 8, 281–338

72 Spanos, N.P. (1994) Multiple identity enactments and multiple personality disorder: a sociocognitive perspective *Psychol. Bull.* 116, 143–165

73 Poole, D.A. *et al.* (1995) Psychotherapy and the recovery of memories of childhood sexual abuse: doctoral-level therapists' beliefs, practices and experiences *J. Consult. Clin. Psychol.* 63, 426–437

74 Dywan, J. (1995) The illusion of familiarity: an alternative to the report-criterion account of hypnotic recall *Int. J. Clin. Exp. Hypnosis* 43, 194–211

75 Ross, D.F. *et al.* (1994) Unconscious transference and lineup identification: toward a memory blending approach, in *Adult Eyewitness Testimony: Current Trends and Developments* (Ross, D.F., Read, J.D. and Toglia, M.P., eds), pp. 80–100, Cambridge University Press

76 Loftus, E.F. (1979) *Eyewitness Testimony*, Harvard University Press

77 Zaragoza, M.S. and Mitchell, K.J. (1996) Repeated exposure to suggestion and the creation of false memories *Psychol. Sci.* 7, 294–300

78 Loftus, E.F. (1997) Memory for a past that never was *Curr. Dir. Psychol. Sci.* 6, 60–65

79 Hyman, I.E., Jr, Husband, T.H. and Billings, F.J. (1995) False memories of childhood experiences *Appl. Cognit. Psychol.* 9, 181–197

80 Schacter, D.L., Kagan, J. and Leichtman, M.D. (1995) True and false memories in children and adults: a cognitive neuroscience perspective *Psychol. Public Policy Law* 1, 411–428