doi:10.1093/brain/awl093 Brain (2006), 129, 1399–1414

# Mechanisms of spontaneous confabulations: a strategic retrieval account

Asaf Gilboa,[1,5] Claude Alain,[1,2] Donald T. Stuss,[1,2,3] Brenda Melo,[4] Sarah Miller[2] and Morris Moscovitch[1,2]

[1]Rotman Research Institute, Baycrest Centre, [2]Department of Psychology, [3]Department of Medicine (Neurology, Rehabilitation Science), University of Toronto, [4]Toronto Rehabilitation Institute, Toronto, Ontario, Canada and [5]Department of Psychology, University of Haifa, Haifa, Israel

Correspondence to: Asaf Gilboa, PhD, Department of Psychology, University of Haifa, Mount Carmel, Haifa 31905, Israel
E-mail: agilboa@psy.haifa.ac.il

**The 'temporality' hypothesis of confabulation posits that confabulations are true memories displaced in time, while the 'strategic retrieval' hypothesis suggests a general retrieval failure of which temporal confusion is a common symptom. Four confabulating patients with rupture of an anterior communicating artery (ACoA) aneurysm, eight non-confabulating ACoA controls and 16 normal controls participated in three experiments designed to test the two hypotheses. In Experiment 1, participants were tested on two continuous recognition tasks, one requiring temporal context distinctions, previously shown to be sensitive to confabulation and another that only requires content distinctions. Both manipulations were sensitive to confabulation, but not specific to it. Temporal context and content confusions (TCCs and CCs) can be explained as failures to make fine-grained distinctions within memory. In Experiment 2, free recall of semantic narratives that require strategic retrieval but are independent of temporal context was used to induce confabulations associated with remote memory, acquired before the onset of amnesia. Confabulators produced significantly more errors. Thus, when retrieval demands are equated, confabulations can be induced in the absence of temporal confusions. Only confabulators conflated semantic content from different remote semantic narratives and introduced idiosyncratic content, suggesting that qualitatively different mechanisms are responsible for distortions due to normal memory failure and for confabulation. Lesion analyses revealed that damage to ventromedial prefrontal cortex is sufficient for temporal context errors to occur, but additional orbitofrontal damage is crucial for spontaneous confabulation. In Experiment 3, we tested whether failure in memory monitoring is crucial for confabulation. Recognition of details from semantic and autobiographical narratives was used to minimize the initiation and search components of strategic retrieval. Only confabulators made more false alarms on both tasks, endorsed even highly implausible lures related to autobiographical events and were indiscriminately confident about their choices. These findings support a strategic retrieval account of confabulation of which monitoring is a critical component. Post-retrieval monitoring has at least two components: one is early, rapid and pre-conscious and the other is conscious and elaborate. Failure of at least the former is necessary and sufficient for confabulation. Other deficits, including TCC and CC, may be required for spontaneous confabulations to arise. The confluence of different sub-components of strategic retrieval would determine the content of confabulation and exacerbate its occurrence.**

**Keywords**: spontaneous confabulations; provoked confabulations; confabulatory amnesia; temporal context confusion; content confusion strategic retrieval; aneurysms

**Abbreviations**: ACoA = anterior communicating artery; CC = content confusion; FOK = feeling of knowing; FOR = feeling of rightness; MTL = medial temporal lobe; PFC = prefrontal cortex; TCC = temporal context confusion

*Received October 31, 2005. Revised February 17, 2006. Accepted March 22, 2006. Advance Access publication April 25, 2006*

## Introduction

Spontaneous confabulations involve statements or actions that reflect unintentional but obvious distortions of memory. Patients who confabulate act upon memories that are obviously false or provide false information without intending to lie, and are unaware of these falsehoods. They sometimes will cling to their false beliefs even when they are aware of contradictory information or when confronted with the truth, giving rise to secondary or provoked confabulations (Kopelman, 1987; Moscovitch, 1989), which are distinct from primary or spontaneous confabulations. It has also been argued that behavioural spontaneous confabulations, which refer to patients who sometimes act upon their false beliefs, are a distinct sub-type of spontaneous confabulation and may be mediated by different mechanisms (Schnider *et al.*, 1996; Schnider, 2003). Whether or not verbal and behavioural confabulations represent different clinical entities is an unresolved issue. Here, we use both Kopelman's and Schnider's definitions to clinically identify and describe patients who confabulate, which will allow for better comparison with other studies of confabulation. We contrast two different accounts of confabulation—one which states that confabulations, particularly spontaneous ones, result from deficits in strategic retrieval against a background of poor memory, and the other which states that confabulations arise from deficits in temporality caused by an inability to suppress currently irrelevant memories and adjust them to ongoing reality.

Confabulation is sometimes found in a sub-group of patients who survived aneurysms of the anterior communicating artery (ACoA; Talland, 1965; Lindqvist and Norlen, 1966; Kapur and Coughlan, 1980; Alexander and Freedman, 1984; Damasio *et al.*, 1985; Vilkki, 1985; Baddeley and Wilson, 1988; Moscovitch, 1989; DeLuca and Cicerone, 1991; DeLuca, 1993; Fischer *et al.*, 1995; Diamond *et al.*, 1997). The 'ACoA syndrome' (Alexander and Freedman, 1984; Damasio *et al.*, 1985; Vilkki, 1985; DeLuca, 1993) includes personality changes, amnesia and confabulations, which may occur independently of each other so that memory deficits may appear in the absence of confabulation (DeLuca, 1993; Diamond *et al.*, 1997). Because the ACoA amnesic syndrome can occur with or without confabulation, the present report focuses on this group of patients to study the specific cognitive and neuroanatomical conditions that produce confabulatory amnesia as compared with amnesia without confabulation.

## Theories of confabulation

There are two primary classes of theories regarding the neurocognitive mechanisms underlying confabulations (for review, *see* Gilboa and Moscovitch, 2002). Temporality accounts suggest that patients have a disturbed sense of chronology so that they can remember the content of events but not their order of occurrence or their relevance to current reality (Talland, 1965; Dalla Barba, 1993; Korsakoff *et al.*, 1996; Schnider *et al.*, 1996; Schnider and Ptak, 1999; Ptak *et al.*, 2001; Schnider, 2003). As a result, patients misattribute aspects of events that occurred at one time to events that occurred at another time or to ongoing reality. A similar account of confabulation suggests that it is a result of the patients' inability to distinguish the source of different memories (source monitoring) or distinguish between real events and imagined ones (reality monitoring; Johnson *et al.*, 1993; Johnson *et al.*, 1997; Johnson and Raye, 1998). Experimental support for a temporality disorder in spontaneous behavioural confabulation comes from a series of studies that used two runs of a continuous recognition paradigm, separated by 1 h (Schnider *et al.*, 1996; Schnider and Ptak, 1999; Schnider *et al.*, 2000). Patients with behavioural spontaneous confabulations showed a specific disproportional deficit on the second run, erroneously choosing stimuli that were relevant on the previous list but not the current one. This was interpreted as reflecting temporal context confusion (TCC) which can lead to conflating of memories from widely dispersed time periods or the inability to suppress currently irrelevant memory cues.

Strategic retrieval hypotheses (Moscovitch, 1989; Burgess and Shallice, 1996; Moscovitch and Melo, 1997; Burgess and McNeil, 1999; Gilboa and Moscovitch, 2002) stem from the observation that confabulation affects remote memories acquired long before brain damage occurred, as much as recent memories that were acquired subsequent to injury. Accordingly, confabulation is considered more a deficit in retrieval rather than in encoding. Strategic retrieval refers to the condition in which the target memory is not elicited directly by the cue but needs to be recovered through a strategic search process akin to problem solving. These strategic retrieval processes operate at input to frame the memory problem and initiate search, constrain it and guide it towards local, proximal cues that can activate associative memory processes. Once a memory is recovered, strategic processes operate at output to monitor the recovered memory and determine if it is consistent with the goals of the memory task and with other knowledge, thereby verifying whether the recovered memory is probably true or false. One way to contrast temporality and strategic retrieval accounts of confabulation is to test confabulation for semantic material. Unlike autobiographical memories, in which confabulations most often occur, semantic memories are independent of temporal context, and any errors found are errors of content (Moscovitch and Melo, 1997).

## The present study

We examine whether temporality accounts of confabulation can be accommodated within the strategic retrieval framework. Spontaneous confabulations by definition defy experimental manipulation, and so like all experimental studies before us, we used paradigms that elicit or provoke confabulation and/or false responses. The relevance and specificity of our findings to spontaneous confabulation arise from the qualitatively distinct patterns of responses exhibited by clinically identified spontaneous confabulators. Because provoked confabulations are defined as normal responses to faulty memory (Kopelman, 1987), qualitatively abnormal responses reflect the mechanisms of spontaneous confabulation in our spontaneous confabulators. In Experiment 1, we administered the standard two runs of the continuous recognition paradigm (Schnider *et al.*, 1996; Schnider and Ptak, 1999) to persons with ACoA amnesia with and without confabulations and to normal controls, to determine whether the deficits observed on this task are related specifically to spontaneous confabulation or more generally to damage to the ventromedial prefrontal cortex (PFC) and/or basal forebrain. We also administered an alternative task that eliminates the temporal component while retaining the requirement for a fine-grained search through memory. Our aim was to demonstrate that the inability to specify the particular temporal context of a memory trace is but one manifestation of a more general failure in memory specification and monitoring, which can affect context as well as content. In Experiment 2, we tested our participants on their memory for common fairy tales and bible stories to explore content versus context confusions in remote memory. Our purpose was to test remote semantic memory, which is context-free, but retain the need for retrieval of complex narratives, which were acquired pre-traumatically, and which characterize confabulations in the personal domain. We categorized errors qualitatively so as to distinguish normal 'gap filling' or provoked confabulations from pathological confabulatory responses. Experiment 3 focused on the proposition that search deficits can exacerbate confabulation and determine their characteristics, whereas monitoring deficits are crucial for their production. Recognition of details from semantic and personal narratives was used in order to minimize the search component and allow a closer examination of monitoring deficits in confabulation. We propose that there is a very early pre-conscious component operating as part of a dual rather than unitary monitoring process, which can explain confabulators' conviction in their false beliefs even when faced with contradictory evidence (for additional discussion of this idea see: Gilboa and Moscovitch, 2002; Schnider, 2003). As a preliminary test of this idea, we included lures for the autobiographical recognition task, which were either plausible details or implausible ones. We hypothesized that whereas normal controls and non-confabulating amnesics would endorse some plausible lures, only confabulators would also endorse implausible lures. The latter finding would suggest that confabulation may arise from a failure of a monitoring mechanism that is qualitatively different from the types of failures seen in controls.

# Methods

## Participants

The present study focused on confabulation following ACoA aneurysm rupture because of the focal nature of the lesion and the obvious patient comparison group (ACoA amnesics with no confabulation), which make this group more appropriate to study than unselected confabulating patients and control amnesics of other aetiologies.

## Confabulating amnesics

Four patients with spontaneous confabulations (GF, SH, NF and GT) participated in the study. GF, SH and GT had sustained ACoA aneurysm ruptures, whereas NF had an ACA-related infarct; however, he is included in the study because of his severe confabulatory syndrome and because his infarct affected some of the regions typically affected by ACoA aneurysms (*see* online Supplementary material for full description of the patients). They all had verbal and behavioural spontaneous confabulations (e.g. packing up to leave the hospital for 'meetings', writing a letter to an obscure personality or demanding treatment for an injury that was never sustained). The patients varied significantly from each other in terms of education and time since injury (Table 1), and had varying degrees of memory impairment and executive functioning deficits (Table 2 and Fig. 1), for which appropriate control patients were matched (*see* below).

## Non-confabulating ACoA controls

Eight patients with ACoA aneurysm rupture and varying degrees of memory loss served as controls. Patients were recruited from the Toronto Rehabilitation Institute, or from the Psychology Department at Baycrest Centre for Geriatric Care. They were well matched for education, time since injury (Table 1) and IQ (Table 2) to the confabulating patients, but were, on average, somewhat younger (although this did not reach statistical significance). Each confabulating patient had at least one ACoA control of the same age (±5 years). Younger and older ACoA controls performed equivalently on the experimental tasks (*see* Results) so that any difference in performance between the groups is unlikely to be the result of age differences.

Table 2 presents the performance of the confabulating and non-confabulating ACoA controls on selected neuropsychological tests. ACoA controls varied with regard to the extent of executive/attentional deficits, some showing average or above average performance, whereas others displayed severe deficits. All ACoA controls showed deficits in at least one memory domain (verbal

**Table 1** Demographics of confabulating patients, ACoA amnesic controls and healthy controls

| | Confabulators ($n$ = 4) | ACoA controls ($n$ = 8) | Healthy controls ($n$ = 16) | Fisher's exact test/ Kruskal–Wallis test |
|---|---|---|---|---|
| Gender (M/F) | 3/1 | 5/3 | 7/9 | $P$ = 0.35; Fisher's exact test |
| Handedness (R/L) | 2/2 | 6/2 | 13/3 | $P$ = 0.71; Fisher's exact test |
| Age (years) mean (SD) (range) | 63.75 (6.05) (57–73) | 51.12 (11.44) (38–68) | 58.87 (10.74) (33–74) | $\chi^2_{(2)} = 3.07$; $P = 0.22$ |
| Education (years) mean (SD) (range) | 15.25 (3.19) (11–20) | 15.75 (2.76) (11–18) | 15.33 (2.72) (11–20) | $\chi^2_{(2)} = 0.714$; $P = 0.7$ |
| Months since loss (month) mean (SD) (range) | 43.5 (40.9) (3–98) | 21.37 (27.44) (4–86) | – | $\chi^2_{(1)} = 0$; $P = 1.0$ |

**Table 2** Neuropsychological performance of confabulators and ACoA controls

| | GF | SH | NF | GT | ACoA controls mean (SD) (range) |
|---|---|---|---|---|---|
| General intellectual | | | | | |
| Shipley verbal (*T*) | 43 | 55 | 38 | 57 | 48.88 (9.51) (34.00–64.00) |
| Abstraction (*T*) | 37 | 62 | 44 | 62 | 55.88 (6.33) (46.00–66.00) |
| Estimated IQ | 92 | 108 | 87 | 116 | 103.25 (10.00) (90.00–119.00) |
| Executive/attentional | | | | | |
| WCST categories | 4 | 3 | 1 | 1 | 3.71 (1.80) (1.00–6.00) |
| Number of errors (*Z*) | −1.44 | −1.86 | −2.94 | −1.81 | −0.51 (0.96) (−1.75–0.87) |
| % Persev. errors (*Z*) | −2.00 | −3.18 | −5.05 | −0.87 | −0.59 (1.32) (−2.7–1.05) |
| Fluency phonemic (*Z*) | −1.88 | −1.67 | −2.41 | −1.65 | −0.97 (1.59) (−2.93–1.22) |
| Semantic (*Z*) | −1.73 | −0.85 | −2.05 | 0.12 | −0.84 (1.45) (−2.76–0.94) |
| SDMT Oral (*Z*) | NA | −1.69 | −1.95 | −0.48 | −0.29 (0.83) (−1.65–0.68) |
| Written (*Z*) | NA | −1.22 | −2.15 | −0.64 | 0.09 (1.20) (−1.89–1.37) |
| TMT Part A (*Z*) | −2.33 | −1.86 | NA | −1.64 | −0.87 (1.31) (−2.33–0.85) |
| Part B (*Z*) | −2.37 | −1.98 | NA | −2.33 | −1.38 (1.58) (−2.88–0.62) |
| Memory | | | | | |
| WMS III LM1(SS) | 4 | 6 | 4 | 4 | 7.25 (2.31) (4.00–11.00) |
| LM2 (SS) | 3 | 7 | 5 | 8 | 7.13 (2.70) (4.00–13.00) |
| VPA1 (SS) | 5 | 5 | 4 | 4 | 5.75 (3.69) (1.00–12.00) |
| VPA2 (SS) | 6 | 6 | 5 | 6 | 6.50 (3.34) (3.00–13.00) |
| Recognition (SS) | 5 | 7 | 6 | 3 | 8.25 (3.11) (5.00–14.00) |
| CFT immediate (*Z*) | −1.64 | −0.99 | −1.34 | −2.33 | −1.4 (0.87) (−2.75–0.6) |
| Delayed (*Z*) | −2.05 | −1.34 | −1.34 | −2.05 | −1.7 (0.79) (−2.75–0.92) |
| RAVLT trial 1 | 3 | 4 | 3 | 5 | 4.75 (1.28) (3.00–7.00) |
| Total learning | 19 | 28 | 20 | 33 | 32.25 (11.15) (22.00–54.00) |
| Delayed | 0 | 5 | 5 | 0 | 4.13 (3.98) (0.00–11.00) |
| Delayed recognition (hits–false alarms) | 5 | 6 | 8 | 6 | 9.75 (2.82) (7.00–14.00) |

CFT = Complex Figure Test; Persev. = perseverative; TMT = Trail-Making Test; LM = logical memory; RAVLT = Rey's Auditory Verbal Learning Test; SDMT = Symbol Digit Modalities Test; SS = scaled score; VPA = WCST = Wisconsin Card Sorting Test; WMS III = Wechsler's Memory Scale III.

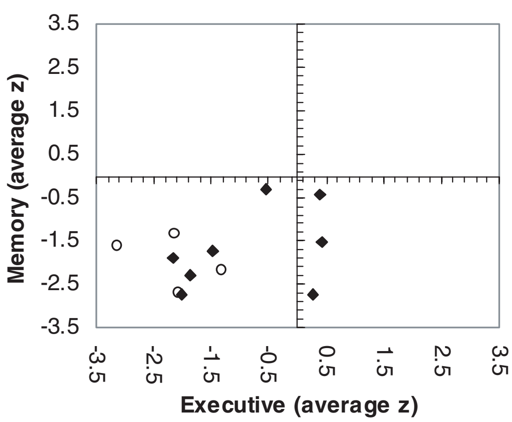


**Fig. 1** Composite executive and memory scores of the confabulating patients (open circles) and non-confabulating ACoA controls (black diamonds).

or visual), as reflected by scaled scores <5 or *Z*-scores lower than −1.64. Generally, confabulating patients' scores fell within the lower range of the scores of non-confabulating controls or below that range. To further explore possible significant differences between confabulators and AcoA controls on general executive and memory functioning, a 'composite executive score' and a 'composite memory score' were computed for each of our patients. The former was the average *Z*-score on the Symbol Digit Modalities Test (oral), Wisconsin Card Sorting Test (% perseverative responses), phonemic verbal fluency and Trails test, part B. The memory score was the average *Z*-score on delayed Logical Memory, delayed Paired Associates from the WMS III, delayed recall on the Rey Auditory Verbal Learning Test and delayed recall on the Rey Complex Figure Test (Fig. 1). Confabulating patients were clearly in the lower range of both memory and executive functioning. It is also clear, however, that they could not be distinguished from the control patients on the basis of their performance on these tests, as four of the control patients performed just as poorly as the confabulators in both domains. The only possible exception is Patient NF who showed particular deficits in executive measures.

## Healthy controls

Sixteen healthy individuals participated as controls in the different studies. They did not differ significantly from the confabulating patients and the ACoA controls with regard to age, education, gender and handedness (Table 1). Their average estimated IQ score was 108 (SD = 11.95; range = 93–133), which also did not differ from that of the other two groups. Not all sixteen controls participated in all of the experiments. The number of controls in each experiment is noted in the Results section.

## Patients with medial temporal lobe (MTL) lesions

Because of the pattern of results obtained in Experiment 1 (*see* below), another group of patients was included in that experiment only. These were four patients with memory impairment resulting from lesions to the MTL. Three of the patients had unilateral MTL excisions due to intractable epilepsy and one had extensive MTL lesions following a traumatic brain injury. These patients were well matched to the other groups in terms of age and education (mean age = 52 years; range = 45–61; mean years of education = 12.5; range = 9–15). Their total learning on the five acquisition trials of the RAVLT was 20, 25, 27 and 30 and their delayed scores were 0, 3, 5 and 9.

## Experiment 1: procedure

The purpose of this experiment was to replicate previously reported deficits of confabulating patients on the second run of two continuous recognition tasks (Schnider *et al.*, 1996). In addition, a third run was constructed, which differed from the other two in that it did not have a temporal component.

### *Temporal context confusion (TCC)*

The same paradigm used by Schnider *et al.* (1996) was used in the present study. Briefly, a series of 120 Snodgrass-like pictures was presented for 2 s with 700 ISI. Eight pictures were repeated 6 times and 72 were presented once only, resulting in 80 new and 40 repeated pictures. Unbeknownst to the subjects, the stimuli were divided into 6 lists of 20 stimuli (8 targets and 12 distracters per list). Subjects were required to press one key for stimuli that had already appeared in the present run and another key for new pictures. After 45 min–1 h, a second run of the same task was performed, using the same stimuli with a different set of eight targets. An increase in the ratio of false-positives to hits on the second versus the first run has been interpreted as representing TCC, reflecting the tendency to respond positively to stimuli on the basis of familiarity from a previous 'reality'. A short practice trial with stimuli that did not appear on the task preceded each of the two runs.

### *Content confusion (CC)*

The second task also consisted of an entirely different set of 120 Snodgrass-like drawings, covertly divided into 6 lists of 20 items. Eight stimuli appeared in all 6 lists, creating a total of 40 targets. The lures for this task were different drawings (exemplars) of particular object types (Fig. 2). There were 20 object types (e.g., trees, chairs, phones, dogs) and each type could have two, three, four or six exemplars. The stimuli were selected on the basis of pilot data with 35 normal controls for adjusting the difficulty of the CC task to the TCC task. Exemplars from the same object type always appeared on consecutive lists. Participants were required to press one key for 'yes' if they had already seen 'exactly the same picture in this series'. A short practice trial with stimuli that did not appear on the list preceded the task. The practice trial also contained different exemplars of the same object, and if participants responded with a 'yes' they were corrected and told that 'this is not the exact same picture as the previous one. They are different horses'. This only occurred twice. The CC task was administered either before or after the two temporal confusion runs, and the order of testing was counterbalanced across subjects. A CC ratio score was calculated using the same baseline as the TCC.

## Lesion analysis

CTs and/or MRIs were obtained for all four confabulators and for seven ACoA controls. A procedure based on Damasio and Damasio

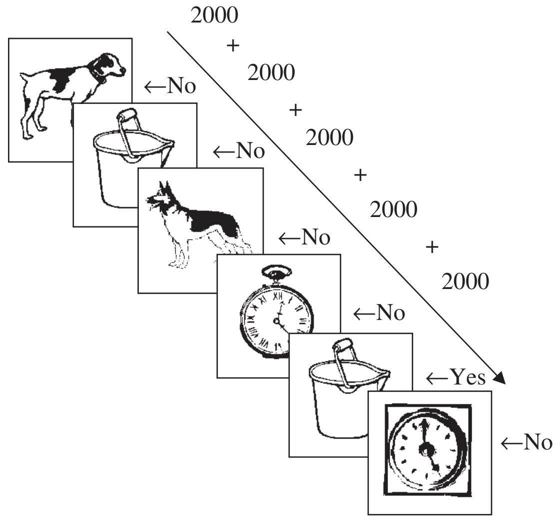


**Fig. 2** Sample of the stimuli of the 'CC' continuous recognition task.

(1989) was used for identifying which frontal lobe regions sustained lesions. However, rather than the standard Brodmann areas, we used the more refined Petrides and Pandya (1994) architectonic divisions for the frontal lobes (Stuss *et al.*, 2002; Stuss *et al.*, 2005). We identified the specific frontal regions that were damaged in each patient (Fig. 3) by superimposing their individual scans on a brain template. Lesions were drawn by the primary author, and for verification and replication, they were independently drawn by D.T.S. and Dr Michael Alexander from Harvard Medical School, Boston, MA, USA. Where there were differences, the latter's delineation of the lesion was used. To determine the relationship between behaviour and lesion, test scores of patients with lesions to a particular region (>25% affected) were compared with the scores of patients without lesions to that region, using non-parametric Mann–Whitney *U*-tests. Only areas that were damaged in three or more patients were evaluated for the brain–behaviour analysis. To determine whether the lesions to regions that appeared to affect test scores were independent of each other (i.e. dissociated) or not, we grouped the cortical architectonic areas into intermediate superordinate anatomical categories adapted from the frontal lobe divisions of Stuss *et al.* (2002, 2005). For the present analysis the relevant regions were orbitofrontal (areas 11, 14, 13, 47/12), ventromedial (area 14), inferior medial cingulate (areas 25 and 24i) and inferior medial paracingulate (area 32). Lesions to other regions did not appear to contribute to deficits on any of the tasks. We then determined the conditional probability that a patient who had damage in one of the defined regions had damage in another. Two regions were classified as being associated if the sum of the two squared conditional probabilities was >0.25, and dissociated if the sum was <0.25 (Stuss *et al.*, 2005).

## Experiment 1: results

The TCC scores of the confabulating patients ranged from 0.36 to 1.01 (Fig. 4), within the range of other confabulators' scores on this task (Schnider *et al.*, 1996). Healthy controls also performed within the expected range, with the highest TCC score being 0.26. Unexpectedly, however, six ACoA controls' scores were as high as those of the confabulating patients, in sharp contrast to previously

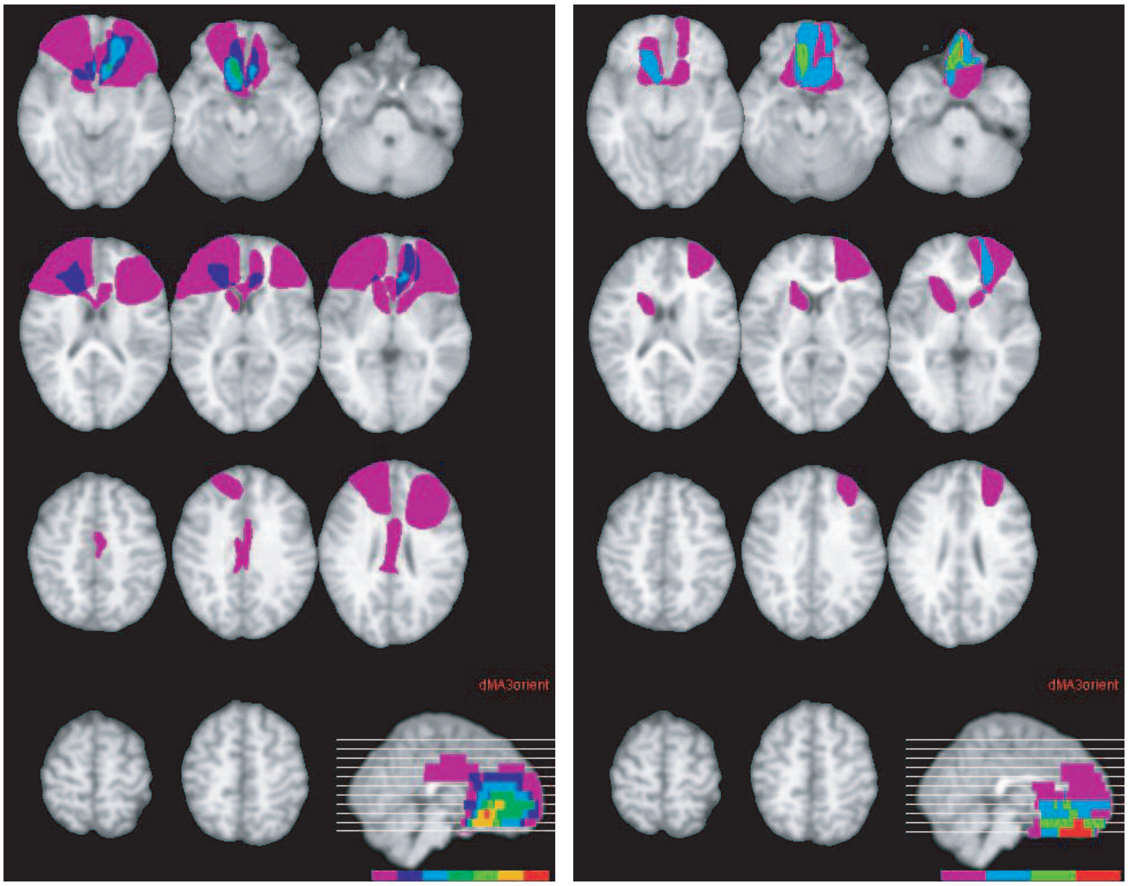


**Fig. 3** Lesion overlap for the ACoA controls (*n* = 7; left) and confabulating patients (*n* = 4; right) superimposed on a template MRI scan. Colour bars represent the number of patients with lesions that overlap a particular region, with purple being one patient and red being all patients (four in the confabulation group and seven in the ACoA control group). Image was created using MRIcro software (Chris Rorden; www.psychology.nottingham.ac.uk/staff/cr1/mricro.html).

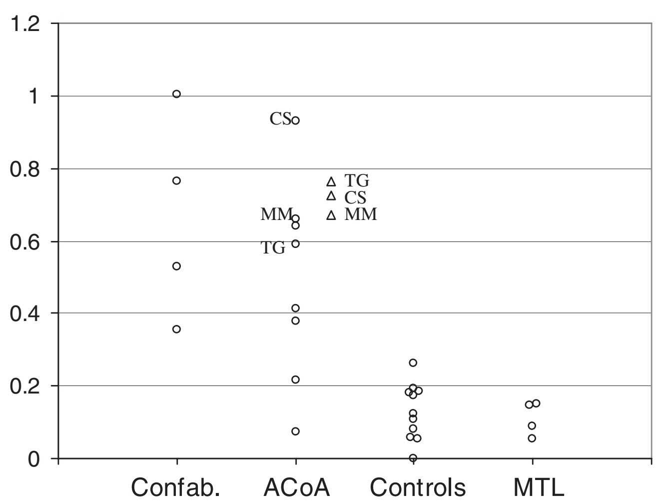


**Fig. 4** Scatter distributions of TCC scores for confabulating ACoA patients, non-confabulating ACoA patients, healthy controls and MTL amnesics. TCC = (FP2/Hits2) − (FP1/Hits1), where FP2 and FP1 are false-positives in Runs 2 and 1, respectively, and Hits2 and Hits1 are the hits on Runs 2 and 1, respectively. ACoA controls who were re-tested (*see* text) are denoted by initials and the results of re-testing are indicated by triangles.

reported scores of non-confabulating amnesics of mixed aetiologies who never exceeded a score of 0.3.

To test whether this high TCC score was due to the different patient control group used in the present study, we administered the test to four male patients with MTL-related memory impairments. One of the patients had a very low hit rate, so his TCC score should be viewed cautiously. The other three had hit rates and TCC scores within the range of the other groups (Table 3). It was also suggested to us that the high TCC scores of non-confabulating ACoA amnesics might arise because the response keys were pressed by our participants rather than by the test administrator (Dr Armin Schnider, personal communication). We re-administered the test to three of our ACoA controls and one ACoA confabulator with verbal rather than motor responses. The results remained as before (Fig. 4). One of the ACoA controls (CS) took a more liberal approach in the second testing session, but this increased her false alarm rate for both runs, leaving her TCC score essentially unchanged.

Another possibility raised was that our ACoA controls may present with behavioural confabulations without verbal spontaneous confabulations. These types of behaviours sometimes go undetected when verbal spontaneous confabulations (which are more conspicuous) are absent (Dr Armin Schnider, personal communication). We re-interviewed over the phone the family members of four of the six ACoA controls who performed poorly on the task. We also questioned the three patients who did return for testing. One

**Table 3** Average number of hits and false alarms on the three continuous recognition tasks

| | Run 1 | | Run 2 | | Content task | | TCC | CC |
|---|---|---|---|---|---|---|---|---|
| | Hits | FA | Hits | FA | Hits | FA | | |
| Conf. | 34.25 | 7.25 | 29.75 | 26.50 | 30.25 | 25.75 | 0.66 | 0.63 |
| ACoA | 34.75 | 2.88 | 29.63 | 17.00 | 34.00 | 9.88 | 0.49 | 0.21 |
| MTL[a] | 30.67 | 2.67 | 27.67 | 5.00 | 31.33 | 6.00 | 0.11 | 0.11 |
| Controls | 38.18 | 3.18 | 34.82 | 7.27 | 36.55 | 6.00 | 0.13 | 0.08 |

[a]For three patients only. The fourth had significantly lower hit rates.

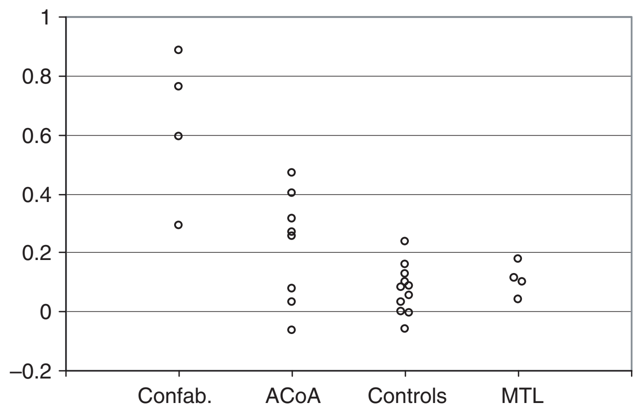


**Fig. 5** Scatter distributions of CC scores for confabulating ACoA patients, non-confabulating ACoA patients, healthy controls and MTL amnesics. CC = (FPcc/Hitscc) – (FP1/Hits1), where FPcc and Hitscc are the number of false-positives and hits for the task, respectively.

patient, who used to confabulate immediately after his aneurysm, reported that, currently (8 years later), he had incidents in which he arrived for a doctor's appointment either very early (several hours) or arrived again a day after he was seen. Though these could be considered behavioural confabulations, they are hard to distinguish from actions based on faulty memory, which are typical consequences of amnesia. Interestingly, this patient also testified that he occasionally had a vivid memory that he knew was not true, so he may well be considered a 'latent confabulator'. For the other patients no behavioural confabulations were reported even when family members were thoroughly questioned, regardless of their history of confabulation. Although it is difficult to rule out the possibility of some residual confabulation in the ACoA controls, this is an unlikely explanation of the results obtained.

To test the alternative hypothesis, namely that TCC reflects a more pervasive strategic memory deficit, we administered the alternative 'CC' task. As can be seen in Fig. 5 and Table 3, CC scores for the confabulators and control groups were very similar to the TCC scores. For the ACoA group, CC scores were lower than the TCC score, even though the same baseline was used for calculating both ratios. This was the result of both a decrease in false alarms and an increase in hits for the content task compared with Run 2 of the TCC task. Nonetheless, some ACoA controls scored considerably higher than any of the healthy controls and MTL patients, and there was considerable overlap in the range of scores of ACoA confabulators and non-confabulators.

### *Neuroanatomy of TCC and CC*

Exploratory analysis of the brain regions that are crucial for performance on the two tasks was performed by comparing the scores of all the patients who had lesions to an area with the scores of patients without damage to that area using non-parametric Mann–Whitney *U*-tests. We considered all areas with a one-tailed $P < 0.05$ where patients with damage had worse scores as critical areas in the processes that determined the measurement (Stuss *et al.*, 2002, 2005).

Regions critically involved in increased TCC were right area 25 (eight patients had lesions to this region; Mann–Whitney $U = 7$) and right ventral (but not orbitofrontal) area 14 (four patients had lesions to that region; Mann–Whitney $U = 6$; Fig. 6A). In contrast, CC resulted from lesions to left inferior area 24 (three patients; Mann–Whitney $U = 4$), inferior area 32 (four patients; Mann–Whitney $U = 6$), orbitofrontal area 14 (four patients; Mann–Whitney $U = 4$) and right orbitofrontal area 11 (three patients; Mann–Whitney $U = 5$; Fig. 6B). Thus, although confabulating patients performed poorly on both tasks, as did some of the ACoA controls, poor performance on these tasks may result from lesions to different cortical regions. However, as might be expected, conditional probabilities indicated that lesions to these regions (i.e. orbitofrontal and ventromedial surfaces) were not independent of each other. Thus, the relationship between the cognitive deficits on these tasks and the lesions leading to these deficits is more consistent with a model of association rather than a model of dissociation (*see* Discussion).

## Experiment 1: discussion

There are two major findings in this study—one functional and the other neuroanatomical. Functionally, we have shown that TCC may be a necessary, but not a sufficient, condition for confabulation. As previously described, this measure was found to be highly sensitive to confabulation (Schnider *et al.*, 1996; 2000; Schnider and Ptak, 1999). All four confabulators in the present study showed elevated TCC. However, elevated TCC in some non-confabulating ACoA amnesics suggests that the specificity of TCC to confabulation needs to be re-examined. Healthy controls and MTL patients had normal TCC scores, suggesting, instead, that TCC is a feature of some patients with ACoA amnesia rather than of confabulation *per se*.

The CC task (which is free of temporal context) produced very similar results to TCC. All of the confabulators had elevated CC scores as did some of the ACoA amnesic controls. Healthy controls and MTL patients were not susceptible to the content manipulation. Thus, behaviourally, we demonstrated that TCC may be one manifestation of a general failure to make fine-grained distinctions within memory. These distinctions can be time-based (i.e. whether an item occurred recently or an hour ago) or content-based (i.e. whether this is the exact same item or only a similar one).

The neuroanatomical analysis pointed to ventromedial cortices (areas 14 and 25) as regions that are crucial for producing TCC, in agreement with previous research (Schnider, 2003). Elevated CC scores were related to ventromedial (inferior areas 24/32) and orbitofrontal areas 14 and 11 and were not dissociable from lesions to regions 14 and 25. The different lesion patterns leading to deficits on the two tasks reflect the fact that some ACoA amnesics were more susceptible to TCC, whereas others were more susceptible to CC. Confabulators, who failed on both tasks, may share neurocognitive characteristics with both groups of patients. A combination or association of deficits, perhaps with additional neurocognitive

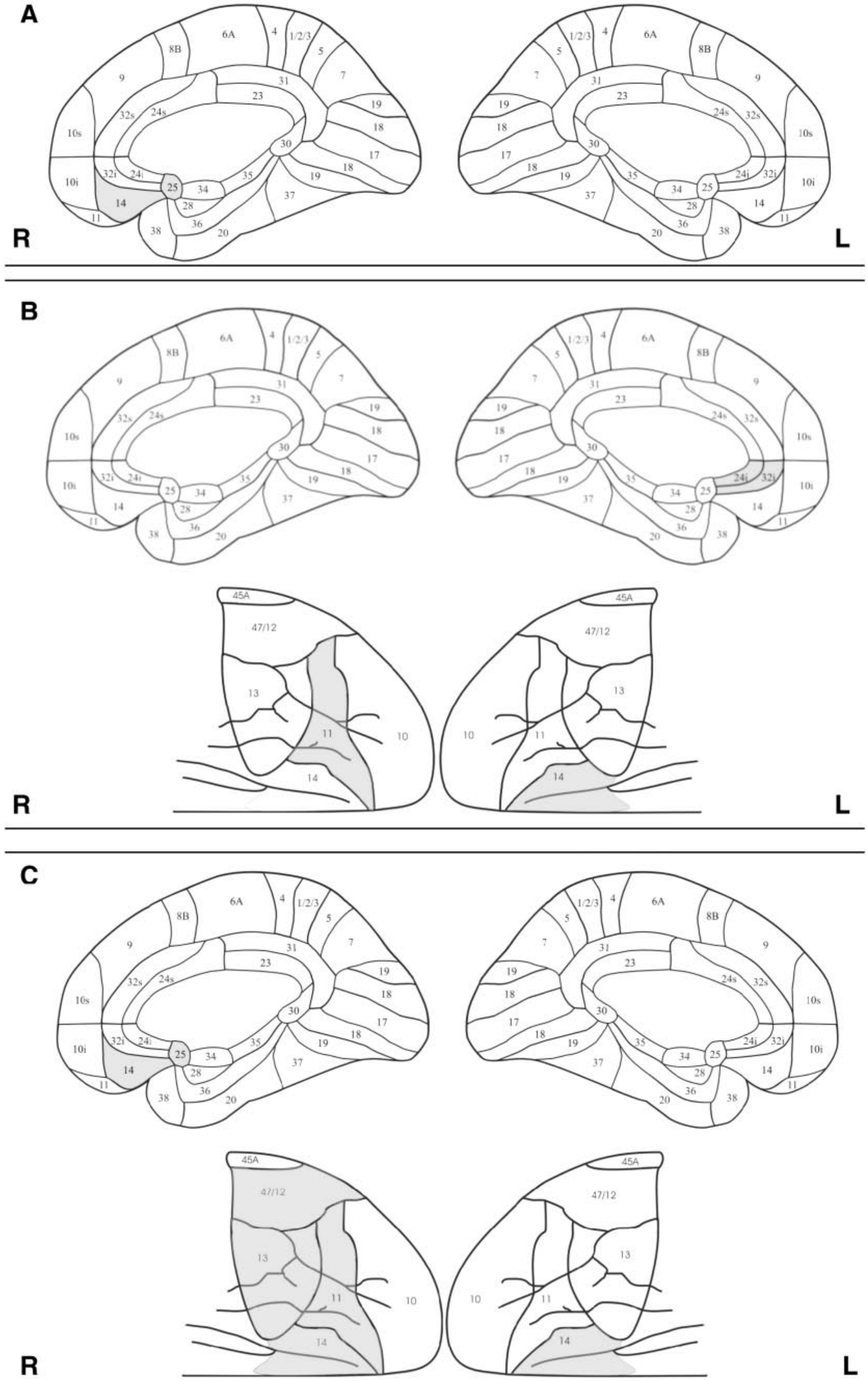


**Fig. 6** Schematic representation of Petrides and Pandya (1994) cytoarchitectonic regions. Regions (only mid-sagittal and orbitofrontal regions) are shown. Shaded regions represent regions that produce significant deficits in TCC (**A**), CC (**B**) and idiosyncratic errors (**C**).

**Table 4** Examples of each type of error on the semantic narratives task

| Confabulators | Distortions of true details | Incorporation of details from another story | Idiosyncratic incorporation |
|---|---|---|---|
| GF | The witch cooks children into gingerbread (Hansel and Gretel) | And god sent Noah to warn [the bad people] but Noah didn't want to go (Noah's Ark with elements of Jonah) | Noah didn't have time to finish the boat when the flood started so he had to get help from somebody (Noah's Ark) |
| SH | The snake guarded the apple. The idea was to get away from the snake and consume the apple (Adam and Eve) | Hansel and Gretel... they walked up the hill to get a pail of water filled up... and Gretel had beautiful long blond hair (Hansel and Gretel confused with Jack and Jill) | He was a passenger on this ship... so he had a beanstalk growing while he was on his way over to Europe (Jack and the Beanstalk) |
| NF | The witch built a gingerbread house and took it to Hansel and Gretel. (Hansel and Gretel) | I think she had two stepsisters and they were mean to her (Snow White with elements from Cinderella) | Should be a chicken... she was a hen I presume... I think she was a leader of a group. They're hunting for something in the woods (LRRH) |
| GT | And she met the big bad wolf and the wolf was dressed like a lady (LRRH) | They were the first two human beings in the world and they had two sons—Jacob and Esau—and I think Jacob was nice and Esau was a mean guy (Adam and Eve with elements from Jacob's life) | Hansel and Gretel lived with their grandmother. And they took a basket of food to the field for lunch for the people who were growing the wheat (Hansel and Gretel) |

characteristics, contributes to confabulation but is not sufficient to produce the full range of symptoms associated with the confabulatory syndrome, as further discussed below.

Experiment 1 suggests that confabulators cannot be distinguished from non-confabulators on the basis of TCC alone. One problem with using continuous recognition paradigms to assess the nature of confabulation is the heavy reliance on false-positives to recently encoded single items as the diagnostic measure. Confabulations are cognitively complex and are associated as much, if not more, with memories (usually autobiographical) encoded long before the injury than with recent memories.

In Experiment 2, we asked participants to retrieve complex semantic narratives that were encoded many years before the injury (Delbecq-Derouesn' *et al.*, 1990). Unlike autobiographical confabulations, which by definition are linked to specific time periods in one's life, errors in the semantic domain are always content errors. Complex narratives also allow for qualitative analyses of errors, which may be more appropriate for characterizing spontaneous confabulations, than a quantitative analysis of false alarms (Kopelman, 1987).

## Experiment 2: procedure

Participants were asked to tell four fairy tales and bible stories from beginning to end with as much detail as possible. These were chosen from those most highly rated on a familiarity Likert scale of 1 (vaguely familiar) to 5 (very familiar). The titles were *Little Red Riding Hood (LRRH)*, *Snow White*, *Jack and the Beanstalk*, *Hansel and Gretel*, *Moses and the Exodus*, *Noah's Ark*, *Adam and Eve in the Garden of Eden*. For patients, stories were recorded and later transcribed; for healthy controls the test administrator wrote down the responses verbatim. General prompting was used to encourage recall of as many details as possible ('can you remember any more?'). If despite their high rating participants could not recall any detail of the story, a set of pre-determined cues, one per story, were used as prompts (e.g. 'If I tell you the word 'giant' can you recall anything from the story?'). Narratives were scored by counting the total number of details recalled for each of the stories. Details were classified as true details, errors and repetitions. Errors were further categorized into (i) distortions of true details; (ii) details from another fairy tale or bible story (including other stories from the list); (iii) idiosyncratic error (any error that could not be identified as one of the previous two categories); (iv) self-corrected errors. The latter were not included in the total sum of errors, because participants had recognized the error they made. Table 4 presents some examples of each type of error. A fifth category of 'Other' was used for classifying utterances that were not related to the story but rather conveyed meta-cognitive comments of different kinds (e.g. 'these stories are really not for children'). Of a total of 76 protocols, 32 were scored by two independent raters. Inter-rater agreement ranged from 98% for true details belonging to the original story to 86% for details that were inserted from another fairy tale or bible story. The latter lower inter-rater agreement was mainly the result of one of the judges not being familiar with a particular story; however, even that rate of agreement was high.

## Experiment 2: results

Confabulators' and ACoA controls' mean familiarity ratings were slightly lower than those of controls (3.75, 3.52 and 4.36, respectively), but the difference was not statistically significant (Kruskal–Wallis $\chi^2_{(2)} = 4.04$; $P = 0.13$). There was no difference between the groups in the average total number of details (Kruskal–Wallis $\chi^2_{(2)} = 3.49$; $P = 0.17$; Fig. 7). There was a significant difference, however, in the overall percentage of errors (Kruskal–Wallis $\chi^2_{(2)} = 8.73$; $P = 0.01$; Fig. 8). Confabulators made significantly more errors than the ACoA controls (Mann–Whitney $U = 0$; $P = 0.007$) and the normal controls (Mann–Whitney $U = 1$; $P = 0.01$). ACoA and normal controls did not differ from each other (Mann–Whitney $U = 22.5$; $P = 0.52$). Patient GT produced fewer errors than the other three patients. His total number of errors was similar to that of three of the healthy controls (Fig. 8). He was generally cautious on this task, producing relatively few details, commenting on his poor memory.

More dramatic differences were noted with regard to the types of errors. There was only a marginal difference between the three groups in the number of distortions of details (Kruskal–Wallis $\chi^2_{(2)} = 5.32$; $P = 0.07$). However, there were significant differences with regard to the number of incorporation of details from another

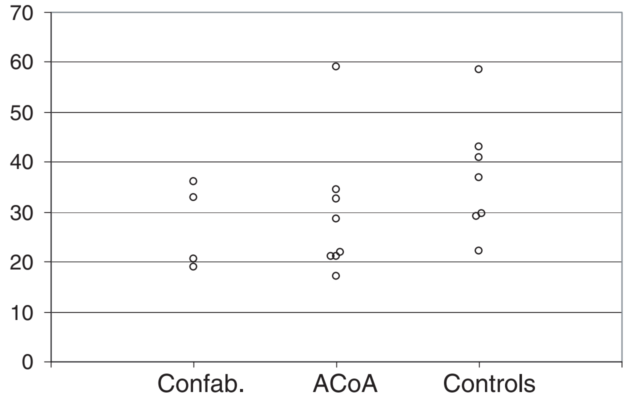


**Fig. 7** Mean number of details produced by participants when recounting semantic narrative.

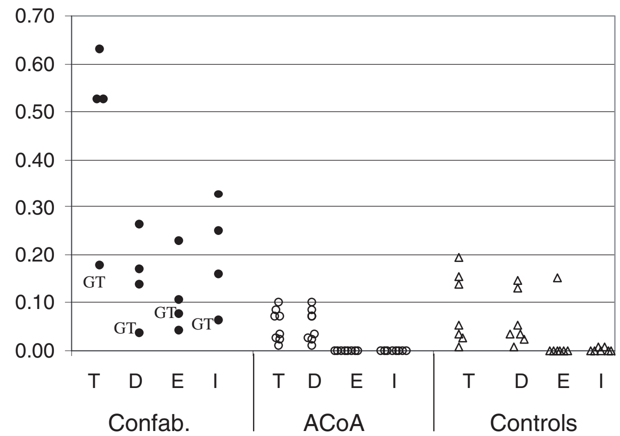


**Fig. 8** Proportion of errors out of the total number of details by error type. T = total number of errors; D = distortions; E = external details (other stories); I = idiosyncratic errors. Patient GT's scores are denoted by initials.

story (Kruskal–Wallis $\chi^2_{(2)}$ =12.57; $P$ = 0.002) and incorporations of idiosyncratic details (Kruskal–Wallis $\chi^2_{(2)}$ =14.02; $P$ = 0.002). In both these categories, the confabulators had significantly more errors. These types of errors were virtually absent in both ACoA and healthy controls. One healthy control erroneously recounted 'Cinderella' instead of 'Snow White'. Later, she noted that she had made an error. Contrary to confabulators, she did not combine details from different stories, but simply recounted a completely different story. Idiosyncratic incorporations were only found in significant numbers in confabulators. Thus, even though differences in total number of details were not strikingly different among the groups, there were important qualitative differences. These qualitative differences were also apparent in Patient GT's pattern of errors (Fig. 8), who despite having fewer errors overall showed increased external details and idiosyncratic ones.

### *Neuroanatomy of semantic confabulations*

We also performed lesion analysis using the percentage score of idiosyncratic incorporations in the fairy tales (Fig. 6C). Interestingly, four of the six regions identified as critically involved in TCC and CC also appeared for this measure of confabulation, namely right ventromedial areas 25 (Mann–Whitney $U$ = 6) and 14 (Mann–Whitney $U$ = 5) as well as left orbitofrontal area 14 (Mann–Whitney $U$ = 6) and right orbitofrontal area 11 (Mann–Whitney $U$ = 3). Two additional regions were unique to this condition, namely right orbitofrontal 47/12 (four patients; Mann–Whitney $U$ = 5) and right orbitofrontal area 14 (seven patients; Mann–Whitney $U$ = 6; Fig. 6C). Note that orbitofrontal 47/12 is different from ventrolateral 47/12 according to the classification of regions we used (Petrides and Pandya, 1994; Stuss *et al.*, 2002, 2005). The former has more extensive connections with limbic and paralimbic MTL structures, playing a central role in memory functions (Petrides and Pandya, 2002). Thus, while lesions to the regions that produce higher TCC and CC scores may contribute and may even be critical for production of confabulation, they are not sufficient for it. Additional lesions, particularly in the orbitofrontal cortex, are necessary for spontaneous confabulation to occur in the sample of ACoA patients studied here.

Because only confabulators had significant numbers of idiosyncratic errors, the regions in this analysis also include the regions that were unique to the four confabulating patients, namely the orbitofrontal cortex (Fig. 3). To determine whether confabulations (and idiosyncratic errors) were not simply the result of larger lesions, we compared the extent of the lesions of confabulators with that of non-confabulators. The percentage of lesion area out of the total brain template was calculated for all patients and the ratios compared across groups. There was no difference between confabulators (mean = 1.98%; range = 0.87–3.45%) and non-confabulators (mean = 2.34%; range = 0.33–6.48%) with regard to lesion size (Mann–Whitney $U$ = 12; $P$ = 0.7).

## Experiment 2: discussion

Confabulations most commonly occur in autobiographical memory, where content and context cannot be teased apart. Errors in the semantic domain are always content errors, and using semantic memory can help address the question of temporality in retrograde memory. The present study suggests that whenever strategic retrieval processes are invoked (e.g. in retrieval of narratives as opposed to facts), confabulators are susceptible to committing errors. Importantly, familiarity was controlled using highly familiar stories and familiarity ratings that did not differ across groups.

Qualitative analyses of error types revealed that both ACoA controls and healthy controls had internal distortions of story details, suggesting that these kinds of errors are common errors associated with memory failure. The confabulators' higher rate of distortions may be viewed as qualitatively similar to 'gap filling' errors and different only in extent. In contrast, confabulators tended to merge details from other stories and produce idiosyncratic details, phenomena that did not occur in the two control groups. Qualitative differences in the kinds of errors produced map onto the distinction between provoked and spontaneous confabulations proposed by Kopelman (1987). According to Kopelman, the former represent a common response to faulty memory, whereas the latter are related to an amnesic syndrome superimposed on a frontal dysfunction. The task we used falls within the realm of the formal definition of provoked confabulations, namely confabulations produced in response to a challenge of memory. Although defined by the way they are tested, the term 'provoked confabulations' was aimed at denoting, more generally, erroneous memories that reflect the typical reconstructive nature of memory. These errors do not differ qualitatively between patients and healthy controls when the strength of the memory trace is weak in controls (Kopelman, 1987). Here, we show errors that, far from being typical, are unique to confabulators in the semantic domain, and that resemble spontaneous verbal confabulation in the autobiographical domain. Previous studies have demonstrated confabulations in the semantic domain, using cue words and historical events (Moscovitch and Melo, 1997) or general knowledge questions (Kopelman *et al.*, 1997). The present report supports their conclusion, and offers control for aspects such as low familiarity and provoked confabulations that may have played a larger role in these studies.

Because only confabulators had idiosyncratic errors, the regions that appear in Fig. 6C (*see* also Fig. 3) can be considered as crucial for the development of a confabulatory syndrome in ACoA patients. The involvement of the orbitofrontal and ventromedial surfaces of the frontal lobes is the most conspicuous finding. We have previously hypothesized, on the basis of a review of the literature (Gilboa and Moscovitch, 2002), that lesions to the ventromedial PFC are sufficient to produce confabulation. Recently, a series of unselected frontal lesion patients (Turner *et al.*, 2005) was tested on the confabulation battery of Dalla-Barba (1993). They found that all patients who produced a total number of confabulations outside the normal range had lesions affecting the ventromedial cortex. The present data are consistent with these reports, because the term 'ventromedial PFC' also encompasses the neighbouring medial orbitofrontal cortex. However, we should remain cautious while drawing conclusions from the anatomical data because some of the non-confabulating patients had been confabulating at an earlier time, when they already had the lesion now entering the analysis as non-confabulation-producing. Confabulation is a dynamic process rather than a static condition, a fact that has puzzled researchers and clinicians of confabulation (DeLuca and Cicerone, 1991; Box *et al.*, 1999; Kopelman, 1999; Schnider *et al.*, 2000; Gilboa and Moscovitch, 2002). Neuroanatomical lesion sites might interact with metabolic or other neurochemical changes to produce such dynamics (DeLuca and Cicerone, 1991; Box *et al.*, 1999; Kopelman, 1999). Alternatively, in keeping with the overall hypothesis raised in the present study, other psychological processes may come into play at later stages such that confabulating patients learn to inhibit their confabulatory ideas (Moscovitch, 1995; Gilboa and Moscovitch, 2002), as further elaborated in the General discussion. The precise psychological mechanisms and associated brain lesions that result in transitory as opposed to enduring confabulatory syndrome are beyond the scope of the present study.

Experiment 2 further demonstrated that temporality cannot fully account for confabulations, and that the data from both Experiments 1 and 2 are explained better within the framework of a strategic retrieval hypothesis. Qualitative distinctions suggest that the mechanism leading to confabulation is different from the mechanisms leading to normal errors.

The strategic retrieval hypothesis suggests that monitoring failures are necessary for confabulations to occur, and Experiment 3 tested this hypothesis by minimizing the search components of strategic retrieval. We predicted that despite this manipulation confabulations would still occur if the deficient monitoring hypothesis was correct.

## Experiment 3: procedure

### *Recognition of details from semantic narratives and autobiographical events*

These tasks were designed to examine monitoring processes in confabulation while minimizing other aspects of strategic retrieval, namely formation of a search strategy and search processes. For the semantic task, we used the same fairy tales and bible stories from Experiment 2. Participants were presented with the title of a story on a computer monitor. Forty sentences (half true and half false) that told the story in chronological order were presented one at a time on the screen, and participants were required to press one key for 'yes' and another for 'no'. The title of the story appeared on the screen throughout, as did reminders of the mapping of the response keys. Participants had 10 s to respond to each sentence, after which it disappeared. An examiner noted responses for details where the response was delayed. After each sentence, a confidence rating (high/low) was required, before the next sentence appeared.

The autobiographical recognition paradigm had essentially the same structure. Six significant events (first dates, weddings, memorable trips, etc.) were selected in consultation with family members of the participants (four from the retrograde and two from the anterograde amnesia period). Before testing, it was verified that the participant was familiar with the event. Family members were interviewed on each event, and 40 sentences (20 true and 20 false) describing the event were constructed. For this task, an additional manipulation was used. False sentences were divided into those containing 'plausible' and 'implausible' details (10 of each type for each event). Family members were asked to create implausible sentences such that the detail described either has never happened in a similar context or has never happened at all. Plausible lures, on the other hand, included details that could have happened in the context of the event but did not occur. In an event concerning a Florida vacation, implausible lures may include 'you spent the night in custody for impaired driving', assuming this never happened and plausible lures may include 'you went swimming in the ocean during your Florida vacation' because this has happened in other contexts and is a likely occurrence in a script like this. Confidence ratings followed each of the sentences, as in the semantic version of the task.

## Experiment 3: results

We hypothesized that if confabulations are primarily the result of failure to initiate and conduct an orderly search through memory, their performance on recognition should be considerably improved. On the other hand, if it is primarily a monitoring failure, the same dramatic differences that were observed on free recall should also be observed on recognition.

On fairy tale recognition there was no significant difference in the hit rates of the three groups (Kruskal–Wallis $\chi^2_{(2)} = 2.39$; $P = 0.3$; Fig. 9). The groups differed on false alarm rates (Kruskal–Wallis $\chi^2_{(2)} = 6.94$; $P = 0.03$), with confabulators scoring significantly higher than ACoA controls (Mann–Whitney $U = 0$; $P = 0.03$) and healthy controls (Mann–Whitney $U = 0$; $P = 0.01$). The two control groups did not differ from each other (Mann–Whitney $U = 10.5$; $P = 0.5$).

In contrast, for autobiographical recognition, there was a significant difference in hit rate (Kruskal–Wallis $\chi^2_{(2)} = 10.38$; $P = 0.006$; Fig. 10). Healthy controls had a higher hit rate than both confabulators (Mann–Whitney $U = 0$; $P = 0.02$) and ACoA controls (Mann–Whitney $U = 0$; $P = 0.01$). In addition, ACoA controls had a higher hit rate than confabulators on this task (Mann–Whitney $U = 0$; $P = 0.03$). The false alarm rates of the three groups were different for both plausible lures (Kruskal–Wallis $\chi^2_{(2)} = 6.45$; $P = 0.04$) and implausible lures (Kruskal–Wallis $\chi^2_{(2)} = 6.63$; $P = 0.03$). This was due to confabulators' higher false alarm rates on both types of errors compared with ACoA controls (Mann–Whitney $U = 0$; $P = 0.03$) and compared with healthy controls (Mann–Whitney $U = 0$; $P = 0.02$). The two control groups

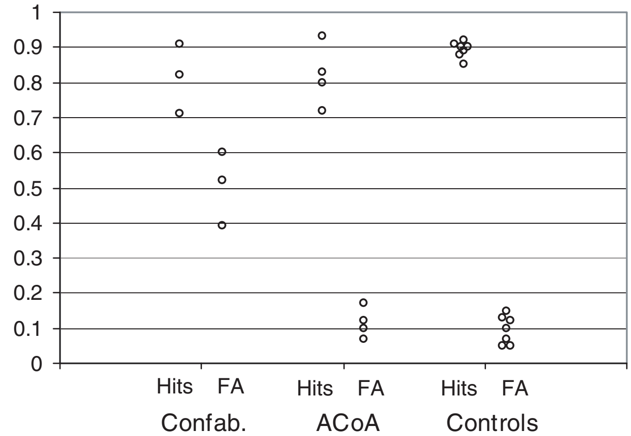


**Fig. 9** Hit and false alarm rates on the recognition of details from semantic narratives.

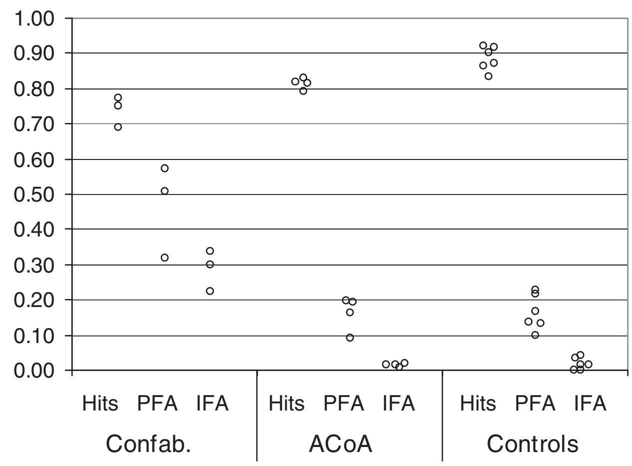


**Fig. 10** Hit and false alarm rates for plausible (PFA) and implausible (IPA) lures on the autobiographical recognition task.

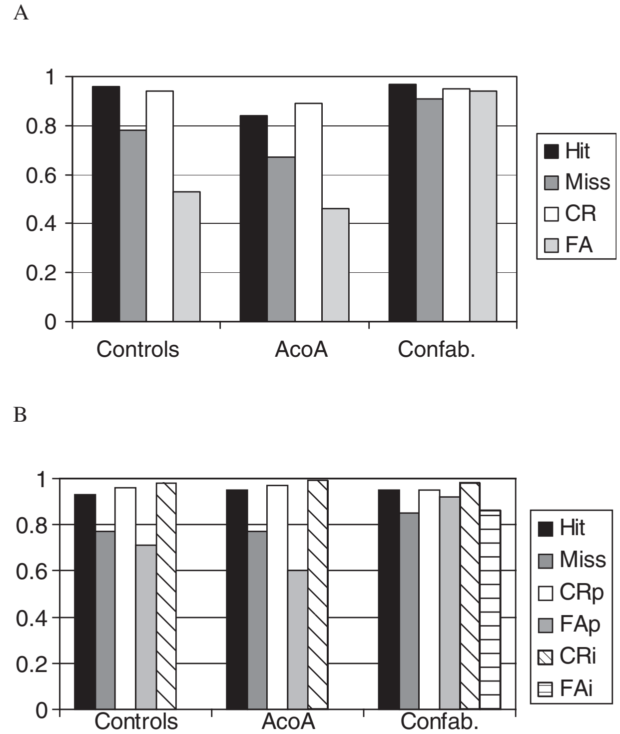


**Fig. 11** Proportion of high confidence responses for semantic (**A**) and autobiographical (**B**) recognition tasks. CR = correct rejection; FA = false alarm; p = plausible; i = implausible.

did not differ from each other on plausible lures (Mann–Whitney $U = 11$; 0.83) or implausible ones (Mann–Whitney $U = 12$; $P = 1.0$).

Thus, unlike the fairy tale recognition task, where confabulators appeared to engage in indiscriminate endorsement of details, for the autobiographical recognition task they had a lower hit rate and a higher false alarm rate. In addition, their failure appeared to be qualitatively different than the errors made by the control groups, who consistently made at least some plausible false alarms but almost never made implausible ones.

### *Confidence ratings*

Following each sentence, participants were required to indicate their confidence in their response as high or low. Overall, participants were quite confident in their responses. Even healthy controls had a surprisingly high percentage of high confidence for misses and false alarms (Fig. 11). This may have occurred because we only provided them with two options, which may not have been sensitive enough to detect subtle differences in confidence. Nonetheless, interesting and similar patterns emerged from both recognition tasks. For the semantic task, both control groups had a lower rate of high confidence responses when they made an error (either miss or false alarm) than when they responded correctly (either hit or correct rejection). ACoA controls were, overall, slightly less confident than healthy controls. In contrast, confabulators had the same high rate of high confidence responses, regardless of accuracy. Thus, they not only indiscriminately endorsed sentences, they did so with high confidence. A similar pattern was seen in the autobiographical recognition task. Confabulators had high rates of confident responses even for implausible false alarms, although it was somewhat lower than that of hits and correct rejections.

## Experiment 3: discussion

### *Recognition of autobiographical and semantic details*

Because strategic retrieval involves multiple processes, it is difficult to decipher the contribution of different processes to confabulation on the basis of free recall paradigms. Here, we used recognition tests of details associated with semantic and personal narratives to minimize the search requirements of the task and focus more on monitoring. On recognition tasks, all the information is presented and one need not establish a search strategy or create retrieval cues to address the long-term memory store. On recognition of semantic details, confabulators had the same hit rate but higher false alarm rate than the control groups. Thus, they demonstrated a bias of indiscriminate endorsement of details. Interestingly, their pattern of confidence ratings was also different from that of the control groups, in that most of their errors, particularly their false alarms, were associated with higher levels of confidence.

The pattern of results for the autobiographical recognition task was somewhat different. Confabulators had a lower hit rate, and higher false alarm rate, suggesting at least part of their performance was not due to a general bias to endorse sentences. The differential false alarm rate between plausible and implausible lures also supports the idea that their performance does not reflect a simple bias. This is also inconsistent with the proposal that confabulation is associated with a general problem with response inhibition. Instead, confabulators appear to commit memory errors, and sometimes even gross ones, even when other aspects of strategic retrieval are minimized. Importantly, these errors were associated with high levels of confidence, again in contrast with controls. On this task, too, a qualitative difference between confabulators and controls was revealed. Both control groups endorsed some of the plausible lures, suggesting that errors made by confabulators for these stimuli may represent exaggerated rates of otherwise normal memory failures (or disagreement between rememberers). In contrast, endorsement of implausible lures was unique to confabulators suggesting a qualitatively different mechanism. As discussed below, we believe this represents a failure of a system that facilitates early, intuitive rejection of false memories that is strongly associated with self-related memories.

## General discussion

Confabulation is complex and multi-faceted, and models that propose a singular underlying mechanism are unlikely to succeed in explaining the variety of its phenomena. Eclectic models of confabulation suggest that a confluence of factors leads to confabulation (Kopelman *et al*., 1997; Johnson and Raye, 1998). Possible contributors include vivid imagination, retrieval failures, source monitoring (Johnson and Raye, 1998), executive dysfunction, cues from the immediate environment and perseverations (Kopelman *et al*., 1997; Mercer *et al*., 1977). In contrast, temporality explanations (Talland, 1965; Dalla Barba, 1993; Korsakoff *et al*., 1996; Schnider *et al*., 1996; Schnider, 2003) and strategic retrieval models (Moscovitch, 1989; Burgess and Shallice, 1996; Moscovitch and Melo, 1997; Gilboa and Moscovitch, 2002) suggest specific underlying cognitive mechanisms that are sometimes broken down into sub-components, but are nonetheless all within the same cognitive domain.

The experiments reported favour strategic retrieval models over the temporality explanations because (i) CCs, in which no temporality was involved were as evident in confabulation as temporal confusions; (ii) confabulations were evident in semantic memory in which specific temporal context related to current reality is not a factor; (iii) confabulations occurred even when initiation and search components of retrieval were minimized, suggesting that defective monitoring at retrieval is a crucial element of confabulation. Anatomically, we found that ventromedial PFC damage always accompanied confabulation, but did not seem to be sufficient for it to occur, as some non-confabulating ACoA patients also had damage to this region. They, too, were impaired on tasks that required fine temporal or content discriminations. Spontaneous confabulations in the present group of ACoA patients depended on additional damage to orbitofrontal cortex, as Schnider and others had speculated.

Below, we elaborate on our previous strategic retrieval account of confabulation (Moscovitch, 1989; Moscovitch and Melo, 1997; Gilboa and Moscovitch, 2002). Specifically, we suggest that failure of retrieval monitoring is crucial for confabulation, but that monitoring is a complex set of processes that includes at least three components. (i) Pre-retrieval feeling of knowing (FOK) (Koriat, 2000; Koriat *et al*., 2000), which may contribute to confabulation by biasing response patterns. Little is known about FOK and confabulation, and hence we will not discuss this further (but *see* Moscovitch, 1989, 1995). (ii) Intuitive immediate post-retrieval feeling of rightness (FOR) (Gilboa and Moscovitch, 2002; Moscovitch and Winocur, 2002; Gilboa, 2004). (iii) Elaborate conscious monitoring of retrieved content for inconsistencies, conflicting evidence and compatibility with the task's requirement (Burgess and Shallice, 1996; Moscovitch and Melo, 1997; Moscovitch and Winocur, 2002). The second aim of this model is to demonstrate how contributing factors such as source-monitoring deficits, perseverations (Kopelman *et al*., 1997; Johnson and Raye, 1998), wishful ideations (Conway and Tacchi, 1996; Fotopoulou *et al*., 2004) as well as temporal confusions (Schnider, 2003) can be integrated into a strategic retrieval model as factors that determine the content of confabulations and exacerbate their production.

## Working with memory (WWM) and the frontal lobes

The basic principles of the model have been described elsewhere (Moscovitch, 1992; Moscovitch and Winocur, 2002) and are presented in Fig. 12. The model is useful for explaining confabulation and its various manifestations.

### *Formulation of a retrieval strategy*

Defective search strategies can lead to no response, as was seen in Patient GT's performance on the fairy tale task (*see* also Moscovithch and Melo, 1997). They can also affect the content of confabulation, leading to associative retrieval, as in the case of perseverated information that is used in response to a current memory task. Confabulating patients sometimes incorporated details from stories they recently recounted into other stories in the fairy tale task and Patient GT retold the same personal story with different characters each time (see Supplementary data). In addition, cues from the immediate environment can serve as responses when an internal search strategy is not applied [*see* online Supplementary information and Kopelman *et al.* (1997) for examples].

### *Specification of retrieval cues*

When a search strategy is formulated and initiated, a set of retrieval cues need to be generated to address the

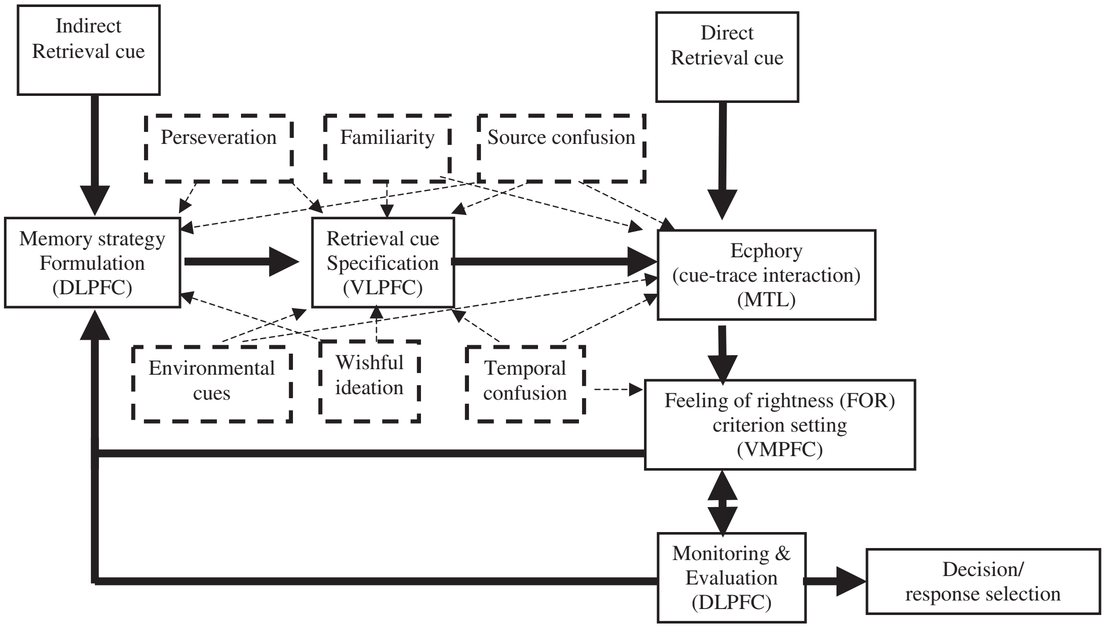


**Fig. 12** WWM and confabulation. When a memory cue does not directly activate a memory representation, reiterative WWM processes need to be invoked in order to generate appropriate memory cues. WWM processes comprise three basic processes: (i) formulation of a search strategy mediated by the DLPFC; (ii) specification of retrieval cues mediated by the ventrolateral PFC (VLPFC); (iii) felt-rightness (VMPFC) and monitoring and evaluation (DLPFC). Cue-dependent retrieval rarely results in confabulation, and when errors occur they are usually considered intrusions or false recognition and are independent of confabulation. Confabulation occurs when the FOR and the monitoring components fail to filter out erroneous memories. The content of confabulation is influenced by processes such as the ones denoted by dashed boxes and arrows that affect WWM processes.

long-term memory store. Faulty formation of cues would lead to erroneous or irrelevant, sometimes haphazard, interactions between them and memory representations ('ecphory'). The content and context confusions from our first experiment are affected by failures to sufficiently specify retrieval cues, responding instead to similar exemplars or to cues from disparate contexts. Such failures can dominate the content of confabulatory phenomena: Burgess and McNeil (1999) demonstrate how generic representations determine the content of confabulation and Dab *et al.* (1999) discuss inappropriate focusing of the temporal dimension. Additionally, wishful ideations (Conway and Tacchi, 1996; Fotopoulou *et al.*, 2004) also serve as powerful generators of irrelevant memory cues, leading to positive biases in the content of confabulation.

## Monitoring

Amnesia, poor search strategies and poor cue specification may lead to activation of inappropriate memories in all ACoA amnesics. Both non-confabulating and confabulating ACoA patients are poor at eliminating aspects of memory that resemble normal memories; however, confabulators also endorse or produce more unusual errors as seen in the idiosyncratic intrusions and the implausible lure conditions in the present study. This is ascribed to failures in monitoring mechanisms.

## Pre-conscious 'feeling of rightness'

One of the most apparent clinical characteristics of confabulators is their absolute conviction in the truthfulness of their erroneous memories even when they can appreciate contradictory evidence and are able to acknowledge the truth; they fail to adjust their FOR. In the present study, confabulators had the same high confidence ratings for false alarms of implausible lures (which controls never endorsed in the first place). Here, we propose a possible mechanism by which FOR might guide memory decisions and affect confabulation, on the basis of three assumptions. (i) FOR is the result of an early categorical selection ('true'/'false') of memory cues based on their compatibility with general cognitive schemata that guide or serve as scaffold for memory reconstruction. (ii) Two factors influence the intensity of FOR: the strength of the schema and the extent of compatibility or deviation of the memory from it. Strong representations endow memory traces with more intense FOR. (iii) The most salient, rich and robust cognitive schema is that of the self (Craik *et al.*, 1999). Compared with other types of information, autobiographical memories evoke an extraordinary sense of confidence in their veracity (Brewer, 1986). One model of autobiographical memory suggests that the self (or 'working self') draws on autobiographical memories as a repository of representations that help it guide behaviour to achieve current goals (Conway and Pleydell-Pearce, 2000). A similar account has also been offered to a related neurological disorder, namely reduplicative paramnesia (Alexander *et al.*, 1979). These authors have coined the term 'affectively burnt in' to explain the difficulty these patients have in overcoming their self-related false beliefs.

When FOR processes break down, the result is acceptance of false memories with high confidence, and given the role of autobiographical memories in guiding behaviour (Conway and Pleydell-Pearce, 2000), behavioural confabulations often arise. These confabulations are often displaced in time; however, this does not mean that the failed mechanism is one of TCC. NF believed that he had broken his collar bone, something that he never experienced and hence is not a temporally displaced confabulation. He acted vigorously upon that belief despite acknowledging inconsistencies (see Supplementary data).

The physiological and neuroanatomical bases of FOR are poorly understood. Our data as well as others' suggest that it crucially depends on the ventromedial and orbitofrontal PFC. One hypothesis regarding the ventromedial PFC is that it acts to integrate cognitive processes with emotional somatic signals to pre-consciously bias decision making (Bechara *et al.*, 1997; 2000). It may accomplish a similar goal in the memory domain, particularly in autobiographical memories that carry emotional meaning, personal significance and sometimes behavioural implications.

### Monitoring and evaluation

Retrieved memories are constantly evaluated for their veracity by comparing them with other retrieved content, available information and with the memory task by processes that are akin to problem-solving procedures. They heavily rely on working memory processes for holding in mind multiple pieces of information, and performing logical operations for deciding whether various retrieved aspects are compatible with each other and with the memory task. They also heavily rely on processes such as conflict detection and conflict resolution. Breakdown of these processes can result in confabulations that not only are inaccurate in the context of retrieval but also lack internal consistency. Believing 'Little Red Riding Hood' was a chicken is wrong in the context of retrieval, but having a hen lead a group through a hunt in the woods lacks internal consistency. SH's memory of a canoe trip involving dozens of patients in their wheel chairs is an example of such discrepancies in the personal domain (see Supplementary data).

### Interaction of monitoring systems

Early, rapid emotionally based decisions to endorse or reject memories may be followed by a more thorough, cognitive assessment of their plausibility. Faced with contradictions, healthy individuals may reluctantly admit a failure of their memory, although the FOR phenomenology may still prevail. Resolution of confabulation may be the result of one system taking over the functions of the other. One of our non-confabulating ACoA controls eloquently described having a vivid memory of going on a hike with his friends. He 'did not trust' his memory because it did not fit with other information, and consequently discovered that the trip did not take place. It may be that some confabulators simply learn not to trust their faulty monitoring system, and learn to rely more heavily on other processes. Long-standing confabulation may involve significant damage to both systems.

### Relation to temporal confusions

The two monitoring systems proposed here parallel or correspond to two forms of representation of temporal information. (i) FOR corresponds to Schnider's concept of TCC as adjustment or suppression or thought to ongoing reality. Both are conceptualized as rapid, automatic and relatively impenetrable to reasoning. Both are directly represented and have a strong affinity to emotion or reward value systems of the brain that have as their epicentre the ventromedial/orbitofrontal cortex. We argue that TCC plays a major role in spontaneous confabulation (behavioural or verbal), and may be necessary for it to occur, but is not sufficient as a single causative mechanism. (ii) The monitoring and evaluation system is related to the constructive nature of autobiographical memory and probably to the dorsolateral PFC (DLPFC). Temporal information in memory is largely inferred during the construction process through explicit cognitive processes rather than directly represented in the brain (Friedman, 1993; Johnson *et al.*, 1993). The breakdown of these strategic constructive processes, and the failure of the explicit monitoring system to detect errors, can result in confabulation with gross chronological misrepresentations (Moscovitch, 1989; Moscovitch and Melo, 1997). This does not necessarily mean that confabulations are misplaced memories. Rather, they are memories that have been constructed erroneously.

## Conclusion

The results of our study support a strategic retrieval account of confabulation. Impairment or failure of one or both monitoring systems, associated with damage to ventromedial and orbitofrontal cortex, appears to be necessary and sufficient for confabulation to occur. Deficits in other processes, however, including impaired memory, cue specification, TCC and CC, may be required for a full confabulatory syndrome to arise. Certainly, the confluence of these different disorders or sub-components or strategic retrieval would determine the content of confabulation and exacerbate its occurrence.

## Supplementary material

Supplementary material is available at *Brain* online.

## Acknowledgements

This research was funded by the Heart and Stroke Foundation of Ontario—Centre for Stroke Recovery and the Louis and

Leah Posluns Centre for Stroke and Cognition and Baycrest and the Canadian Institute for Health Research (CIHR). We would like to thank Dr Armin Schnider for very helpful comments on our results and interpretation and Dr Michael Alexander for drawing the lesions on the CT scans. We would also like to thank Dr Brian Richards for his help with patient recruitment and Dr Darlene Floden for her invaluable help with lesion analysis methods. Finally, we would like to thank the patients and their families for contributing their valuable time to participate in this research.

## References

Alexander MP, Freedman M. Amnesia after anterior communicating artery aneurysm rupture. Neurology 1984; 34: 752–7.

Baddeley A, Wilson B. Frontal amnesia and the dysexecutive syndrome. Brain Cognit 1988; 7: 212–30.

Bechara A, Damasio H, Tranel D, Damasio AR. Deciding advantageously before knowing the advantageous strategy. Science 1997; 275: 1293–5.

Bechara A, Damasio H, Damasio AR. Emotion, decision making and the orbitofrontal cortex. Cerebral Cortex 2000; 10: 295–307.

Brewer WF. What is autobiographical memory? In: Rubin DC, editor. Autobiographical memory. New York: Cambridge University Press; 1986. p. 25–49.

Burgess PW, McNeil JE. Content-specific confabulation. Cortex 1999; 35: 163–82.

Burgess PW, Shallice T. Confabulation and the control of recollection. Memory 1996; 4: 359–411.

Conway MA, Pleydell-Pearce CW. The construction of autobiographical memories in the self-memory system. Psychol Rev 2000; 107: 261–88.

Conway MA, Tacchi PC, Craik FI, Moroz T, Moscovitch M, Stuss DT, et al. In search of the self, a positron emission tomography study. Psychol Sci 1999; 10: 26–34.

Dab, S., Claes, T., Morais, J., and Shallice, T. (1999). Confabulation with a selective descriptor process impairment. Cognitive Neuropsychology, 16: 215–242

Dalla Barba G. Different patterns of confabulation. Cortex 1993; 29: 567–81.

Damasio H, Damasio AR. Lesion analysis in neuropsychology. New York: Oxford University Press; 1989.

Damasio AR, Graff-Radford NR, Eslinger PJ, Damasio H, Kassell N. Amnesia following basal forebrain lesions. Arch Neurol 1985; 42: 263–71.

Delbecq-Derouesn' J, Beauvois MF, Shallice T. Preserved recall versus impaired recognition. Brain 1990; 113: 1045–54.

DeLuca J. Predicting neurobehavioral patterns following anterior communicating artery aneurysm. Cortex 1993; 29: 639–47.

DeLuca J, Cicerone KD. Confabulation following aneurysm of the anterior communicating artery. Cortex 1991; 27: 417–23.

Diamond BJ, DeLuca J, Kelley SM. Memory and executive functions in amnesic and non-amnesic patients with aneurysms of the anterior communicating artery. Brain 1997; 120: 1015–25.

Fischer RS, Alexander MP, D'Esposito M, Otto R. Neuropsychological and neuroanatomical correlates of confabulation. J Clin Exp Neuropsychol 1995; 17: 20–8.

Fotopoulou A, Solms M, Turnbull OH. Wishful reality distortions in confabulation: a case report. Neuropsychologia 2004; 42: 727–44.

Gilboa A. Autobiographical and episodic memory—one and the same? Evidence from prefrontal activation in neuroimaging studies. Neuropsychologia 2004; 42: 1336–49.

Gilboa A, Moscovitch M. The cognitive neuroscience of confabulation: a review and a model. In: Baddeley AD, Kopelman MD, Wilson BA, editors. Handbook of memory disorders, 2nd edn. London: John Wiley and Sons; 2002. p. 315–42.

Johnson MK, Raye CL. False memories and confabulation. Trends Cogn Sci 1998; 2: 137–45.

Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. Psychol Bull 1993; 114: 3–28.

Johnson MK, O'Connor M, Cantor J. Confabulation, memory deficits, and frontal dysfunction. Brain Cogn 1997; 34: 189–206.

Kapur N, Coughlan AK. Confabulation and frontal lobe dysfunction. J Neurol Neurosurg Psychiatry 1980; 43: 461–3.

Kopelman MD. Two types of confabulation. J Neurol Neurosurg Psychiatry 1987; 50: 1482–7.

Kopelman MD, Ng N, Van den Brouke O. Confabulation extending across episodic memory, personal and general semantic memory. Cogn Neuropsychol 1997; 14: 683–712.

Koriat A. The feeling of knowing: some metatheoretical implications for consciousness and control. Conscious Cogn 2000; 9: 149–71.

Koriat A, Goldsmith M, Pansky A. Toward a psychology of memory accuracy. Annu Rev Psychol 2000; 51: 481–537.

Korsakoff SS, Banks WP, Karam SJ. Medico-psychological study of a memory disorder. 1889. Conscious Cogn 1996; 5: 2–21.

Lindqvist G, Norlen G. Korsakoff's syndrome after operation on ruptured aneurysm of the anterior communicating artery. Acta Psychiatr Scand 1966; 42: 24–34.

Mercer B, Wapner W, Gardner H, Benson DF. A study of confabulation. Arch Neurol 1977; 34: 429–33.

Moscovitch M. Confabulation and the frontal systems: strategic versus associative retrieval in neuropsychological theories of memory. In: Roediger HL, Craik FI, editors. Varieties of memory and consciousness: essays in honor of Endel Tulving. Hillsdale, NJ: Lawrence Erlbaum Associates; 1989. p. 133–60.

Moscovitch M. Memory and working-with-memory: a component process model based on modules and central systems. J Cogn Neurosci 1992; 4: 257–67.

Moscovitch M, Melo B. Strategic retrieval and the frontal lobes: evidence from confabulation and amnesia. Neuropsychologia 1997; 35: 1017–34.

Moscovitch M, Winocur G. The frontal cortex and working with memory. In: Stuss DT, Knight R, editors. Principles of frontal lobe function. New York: Oxford University Press; 2002. p. 188–209.

Petrides M, Pandya DN. Comparative architectonic analysis of the human and the macaque frontal cortex. In: Boller F, Grafman J, editors. Handbook of neuropsychology. Vol. 9. Amsterdam: Elsevier; 1994. p. 17–58.

Petrides M, Pandya DN. Association pathways of the prefrontal cortex and functional observations. In: Stuss DT, Knight RT, editors. Principles of frontal lobe function. New -York: Oxford University Press; 2002. p. 31–50.

Ptak R, Birtoli B, Imboden H, Hauser C, Weis J, Schnider A. Hypothalamic amnesia with spontaneous confabulations: a clinicopathologic study. [See comment]. Neurology 2001; 56: 1597–600.

Schnider A. Spontaneous confabulation and the adaptation of thought to ongoing reality. Nat Rev Neurosci 2003; 4: 662–71.

Schnider A, Ptak R. Spontaneous confabulators fail to suppress currently irrelevant memory traces. [See comment]. Nat Neurosci 1999; 2: 677–81.

Schnider A, von Daniken C, Gutbrod K. The mechanisms of spontaneous and provoked confabulations. Brain 1996; 119: 1365–75.

Schnider A, Ptak R, von Daniken C, Remonda L. Recovery from spontaneous confabulations parallels recovery of temporal confusion in memory. Neurology 2000; 55: 74–83.

Stuss DT, Alexander MP, Floden D, Binns MA, Levine B, McIntosh AR, et al. Fractionation and localization of distinct frontal lobe processes: evidence from focal lesions in humans. In: Stuss DT, Knight RT, editors. Principles of frontal lobe functions. New York: Oxford University Press; 2002. p. 392–407.

Stuss DT, Alexander MP, Shallice T, Picton TW, Binns MA, Macdonald R, et al. Multiple frontal systems controlling response speed. Neuropsychologia 2005; 43: 396–417.

Talland G. Deranged memory. New York: Academic Press; 1965.

Turner M, Cipolotti L, Yousry T, Shallice T. An inferior medial localisation for confabulation: evidence from an unselected frontal series. British Neuropsychological Society Meeting. University of London; 2005.

Vilkki J. Amnesic syndromes after surgery of anterior communicating artery aneurysms. Cortex 1985; 21: 431–44.