CORTEX 87 (2017) 16–30

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special issue: Research report

Ventromedial prefrontal cortex generates pre-stimulus theta coherence desynchronization: A schema instantiation hypothesis

Asaf Gilboa [a,b,c,*] *and Morris Moscovitch* [a,b]

[a] *Rotman Research Institute, Baycrest Centre, Toronto, Ontario, Canada*
[b] *Department of Psychology, University of Toronto, Toronto, Ontario, Canada*
[c] *Toronto Rehabilitation Institute, University Health Network, Toronto, Ontario, Canada*

ARTICLE INFO

Article history:
Received 20 May 2016
Reviewed 13 July 2016
Revised 28 August 2016
Accepted 11 October 2016
Published online 31 October 2016

Keywords:
Ventromedial prefrontal cortex
Memory schema
Confabulation
Face recognition
Theta coherence
Desynchronization

ABSTRACT

The ventral medial prefrontal cortex (vmPFC) has long been implicated in monitoring of memory veracity, and more recently also in memory schema functions. In our model of strategic retrieval the two are related. We have proposed that the vmPFC has two schema-dependent functions: (i) to establish context-relevant templates against which the output of memory systems can be compared; (ii) to mediate automatic decision monitoring processes to ensure that only those responses that meet the criterion are enacted. Electroencephalogram (EEG) data were used to provide evidence that vmPFC supports both functions, and that schema instantiation informs monitoring. Participants viewed pictures of acquaintances, along with those of famous and nonfamous people, and were asked to respond positively only to pictures of individuals they had met (*personal familiarity*). The Self serves as a super-ordinate cognitive schema, facilitating accurate endorsement of acquaintances and exclusion of non-personal but familiar faces. For the present report we focused on pre-cue tonic oscillatory activity. Controls demonstrated theta coherence desynchronization between medial prefrontal areas, inferotemporal and lateral temporal cortices. These oscillatory coherence patterns were significantly reduced in patients with vmPFC damage, especially in those with clinical histories of spontaneous confabulation. Importantly, these pre-stimulus cortico-cortical desynchronizations predicted post-cue automatic memory activation, as indexed by a familiarity modulation of the face-sensitive posterior cortical N170. Pre-cue desynchronization also predicted early post-cue frontal positive modulation (P230) and response accuracy. The data are consistent with a schema instantiation model that suggests the vmPFC biases posterior neocortical long-term memory representations that enhance automatic memory cue processing and informs frontally-mediated rapid memory monitoring (P230). Damage to these structures can lead to inaccurate, context-irrelevant activation of schemas. These, in turn, impair monitoring signals and can lead to confabulation when memory control processes are also deficient.

* *Corresponding author*. Rotman Research Institute, Baycrest Centre, Toronto, Ontario, Canada.
E-mail address: agilboa@research.baycrest.org (A. Gilboa).

http://dx.doi.org/10.1016/j.cortex.2016.10.008
0010-9452/

1. Introduction

Studies of spontaneous confabulation in patients with medial prefrontal cortex (mPFC) lesions have contributed significantly to the understanding that memory is reconstructive, which allows flexibility on the one hand but also leads to inaccuracies (Moscovitch, 1989; Schacter, Norman, & Koutstaal, 1998; Schacter, Guerin, & St. Jacques, 2011). These studies also raise the converse fundamental question: What processes prevent memory from being wildly inaccurate? The answer to this question is complex, and undoubtedly involves various processes mediated by different brain structures, as indicated by the occurrence of confabulation in patients with no obvious damage to the ventral medial prefrontal cortex (vmPFC) such as Alzheimer's disease, traumatic brain injury (TBI), thalamic strokes etc. (Gilboa & Moscovitch, 2002; Gilboa & Verfaellie, 2010; Kopelman, 2010; Schnider, 2008). In this paper, however, we focus on the vmPFC because it is by far the most consistent structure that has been implicated and that figures prominently in neurobiological theories of memory construction.

A variety of theories have been proposed to account for confabulation and many pertain to mPFC function. These include reality monitoring (Schnider, 2008; Schnider, Nahum, & Ptak, 2017), temporality/source confusion theories (Dalla Barba, Brazzarola, Marangoni, Barbera, & Zannoni, 2017; Dalla Barba & La Corte, 2013; Johnson, O'Connor, & Cantor, 1997; Serra et al., 2014), motivational accounts (Fotopoulou, 2010), strategic retrieval models (Burgess & Shallice, 1996; Moscovitch & Melo, 1997; Turner & Coltheart, 2010) and eclectic accounts that suggest cross-domain confluence of factors with emphasis on executive and memory dysfunction (Bajo, Fleminger, Metcalfe, & Kopelman, 2017; Kopelman, Ng, & Van den Brouke, 1997). What these models have in common is that they view mPFC damage as biasing or distorting the way incoming information is interpreted (e.g., reality filtering) or as corrupting the way control processes operate on retrieved information (e.g., source confusion, motivational accounts). Our own model incorporates both aspects, a feature shared by other multi-factorial models (Bajo et al., 2017; Burgess & Shallice, 1996).

In previous work we proposed that the mPFC has two related functions in this regard: The first is to establish templates or schemas which enable setting criteria against which the output of memory systems, such as the hippocampus, can be compared (Gilboa, 2004; Gilboa et al., 2006, Gilboa, Alain, He, Stuss, & Moscovitch, 2009). Put in other words, does the output satisfy the goals of the memory task which are established by contextual relevance (Fig. 1A). In this regard, the template or schema will also influence or guide perception and encoding. The second, related function of the mPFC is to monitor decision processes (Moscovitch & Winocur, 2002) to ensure that only those responses that meet the schema-driven criteria are enacted (Fig. 1B). In much of our previous research, we have emphasized the latter, monitoring aspect, but our most recent work has led us to consider more systematically the role of the mPFC in schema representation and instantiation which operate at encoding as well as retrieval (Ghosh & Gilboa, 2014; Ghosh, Moscovitch, Melo Colella, & Gilboa, 2014). In this paper, we will provide evidence that mPFC supports both functions, and that schema instantiation informs monitoring. Because our conceptions of mPFC functions are derived, in part, from studies of confabulation, we briefly review the relevant literature.

The idea that mPFC was implicated in monitoring emerged from observations of confabulation in which patients provided patently erroneous answers even when they had sufficient relevant information to discount them (Moscovitch, 1989). In more controlled studies, confabulating patients were more likely to endorse 'critical' lures, including bizarre or unusual ones with respect to autobiographical events (Gilboa et al., 2006) and semantic facts (Kan, Larocque, Lafleche, Coslett, & Verfaellie, 2010). Moreover, and consistent with the monitoring view, confabulating patients' confidence in their responses is high and does not vary with accuracy, as it does in controls (Gilboa et al., 2006; Hebscher, Barkan-Abramski, Goldsmith, Aharon-Peretz, & Gilboa, 2015). Monitoring processes, which lead to a "feeling of rightness" (Moscovitch & Winocur, 2002; Gilboa et al., 2006, 2009) are very rapid, occurring within about 200 msec of stimulus onset, and underlie the tendency of confabulating patients to act on their memory beliefs even when they can appreciate contradictory evidence. The output of this monitoring process serves as the basis for memory control. Deficits on both processes, caused by damage to the ventromedial and orbitofrontal region of mPFC, respectively (Fig. 1C), are necessary for confabulation to occur (Hebscher et al., 2015), forming action intentions (Uretzky & Gilboa, 2010) that can also lead to behavioral confabulation (Schnider, 2008).

More recently, studies linking the mPFC to schema formation and instantiation (Brod, Lindenberger, Werkle-Bergner, & Shing, 2015; Ghosh & Gilboa, 2014; Ghosh et al., 2014; Gilboa, 2004, 2010; Preston & Eichenbaum, 2013; van Kesteren, Ruiter, Fernandez, & Henson, 2012; Spalding, Jones, Duff, Tranel, & Warren, 2015) have suggested an alternate, or complementary, interpretation of the causes of confabulation. According to this interpretation, confabulation arises because damage to the mPFC leads to disorders in schema representation and instantiation. Schemas, defined as "adaptable associative networks of knowledge extracted over multiple similar experiences" (Ghosh et al., 2014, p. 12057), may be corrupted, over-extended, or hyper-specified, with possible failure to arbitrate between competing schemas, much in the way that similar effects follow damage to structures affecting semantics (Crutch & Warrington, 2011; Hodges, Salmon, & Butters, 1992; Lambon Ralph, Ehsan, Baker, & Rogers, 2012; Murre, Graham, & Hodges, 2001). An appropriate schema provides the organizational structure for interpreting complex events, for encoding them, for searching for them in memory, and for providing a template against which one can determine whether recovered memories are consistent with the schema that has been instantiated. Thus, confabulating patients' misinterpretation of ongoing events, such as believing one is at work when there is evidence that one is in a hospital (Moscovitch, 1989), may result from activating an inappropriate work schema with the patient left trying to make incongruent information, such as hospital beds and wards, consistent with the instantiated schema. Likewise,

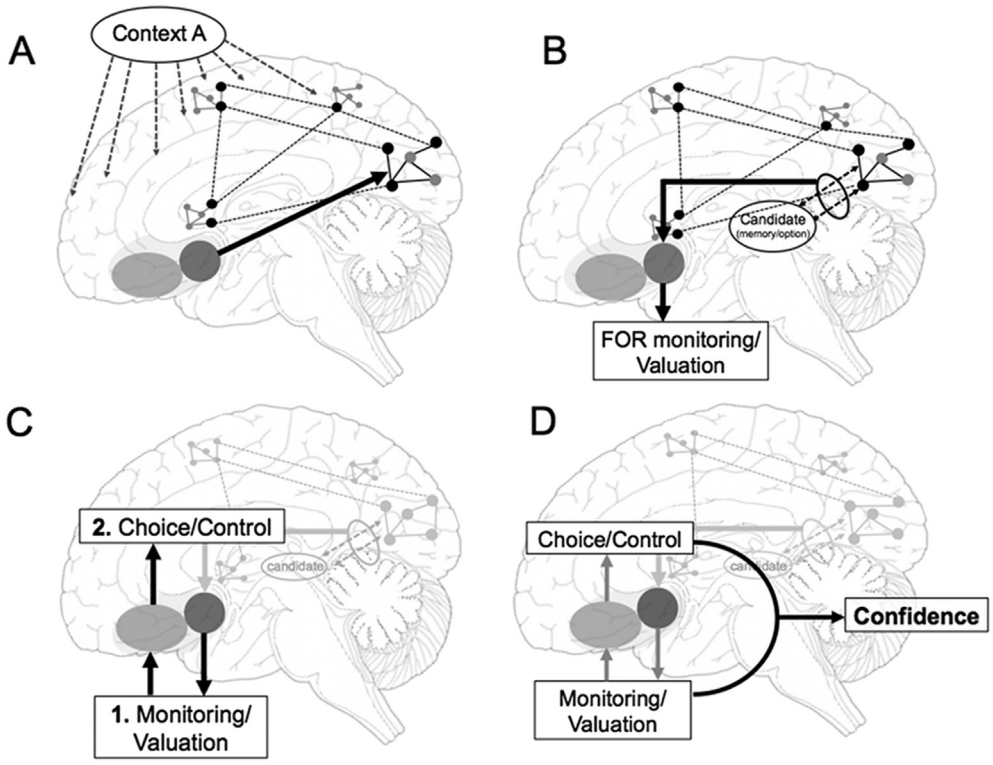

Figure 1 – (A) Schema instantiation: The subcallosal vmPFC biases multimodal schema representations in posterior neocortex based on relevance to the current context. Relevant schema is activated and maintained such that incoming information is interpreted in relation to it. (B) Subjective valuation/Feeling of rightness (FOR) monitoring: Match between activated schema and candidates (memory traces or choice options) automatically drives vmPFC-based FOR/valuation. (C) Memory control and choice behavior: (1) Automatic monitoring and valuation, mediated by subcallosal vmPFC, serves as basis for (2) metacognitive control in memory (decision to report or act on a memory candidate) or choice-behavior in decision-making (choosing one option), mediated by posterior orbitofrontal cortex (OFC). (D) Confidence: Confidence estimates reflect the relationship between first-order monitoring/valuation and second-order control/choice. These relationships are U-shaped and are associated with more extensive anterior region of vmPFC and posterior OFC. This confidence signal informs decisions and actions in both mnemonic and nonmnenomic domains (Reprinted from Hebscher & Gilboa, 2016). The present study focuses on processes (A) and (B).

because schemas in confabulating patients are corrupted, memory search under their guidance often either fails to yield any memory (Moscovitch & Melo, 1997), or yields a memory that matches a corrupted or over-extended schema, leading to confabulations. Others (Attali, De Anna, Dubois, & Dalla Barba, 2009; Burgess & Shallice, 1996; Serra et al., 2014) have also highlighted the role of generic, well-learned information in determining the content of confabulation. Poorly operating schemas can paradoxically also lead to more accurate memory retrieval in conditions in which schemas in healthy individuals lead to memory distortion. Patients with vmPFC lesions endorse fewer critical lures than controls in the Reese-Roediger-McDermott paradigm due to impaired schema-driven extraction of common themes from discrete events (Melo, Winocur, & Moscovitch, 1999; Warren, Jones, Duff, & Tranel, 2014).

As noted above, if disorders in schema representation and instantiation are one cause of confabulation, deficits should be observed even on tasks that do not involve memory directly. To test this hypothesis, Ghosh et al. (2014) had participants instantiate two schemas sequentially, such as 'going to bed' followed by 'a visit to a doctor', and then determine whether words such as 'pajamas' and 'needle' belonged to the instantiated schema. They found that unlike controls, who had no difficulty in this task, vmPFC patients with spontaneous confabulation either tended to endorse inappropriate items or dramatically slowed down their responses, especially for the second schema in the sequence.

Based on this recent evidence that schema related processes are mediated by vmPFC and may be central to spontaneous confabulation, we decided to examine data that had not been analyzed from a previously published study (Gilboa et al., 2009). In that study, we showed participants pictures of acquaintances, along with those of famous and nonfamous people whom the participant had not met, while recording electroencephalograms (EEGs) and analyzing their evoked response potentials (ERPs) to the pictures. Participants were asked to respond positively only to pictures of individuals they had met which we term *personal familiarity*, making it a form of exclusion recognition memory task in which familiar, but non-personal, faces are rejected. In this task, the self serves as a contextual template allowing accurate attribution of

familiarity. In that sense, the self functions as a superordinate schema containing both abstracted knowledge of prior experiences with particular persons and also, possibly, specific instances of encounters, although the latter were less likely to be accessed in our task. This approach to the self as a cognitive schema supporting memory processes is well established in both healthy controls (e.g., Lieberman, Jarcho, & Satpute, 2004; Rameson, Satpute, & Lieberman, 2010; Rogers, Kuiper, & Kirker, 1977) and patient populations (e.g., Segal, 1988), and has been linked to the mPFC in general (e.g., Craik et al., 1999; Kelley et al., 2002) and vmPFC in particular (Lieberman et al., 2004; Rameson et al., 2010; Marquine et al., 2016). Face recognition is a high expertise domain of the kind that generates high self-schema representations and that preferentially engages vmPFC during self-related judgments (Lieberman et al., 2004; Rameson et al., 2010). Moreover, when faces are used to trigger prior knowledge representations (in the form of widespread vmPFC, hippocampal and posterior neocortical co-activations), encoding of new information is enhanced consistent with known effects of schemas on memory formation (Liu, Grady, & Moscovitch, 2016).

Gilboa et al. (2009) found that patients with vmPFC damage were less accurate than controls in discriminating personally familiar from unfamiliar faces. This difference in performance was also notable at the level of the ERP. In controls, the amplitude of the N170, a face-sensitive ERP component, was greater for personally familiar faces and famous people than for unfamiliar faces. Even faces of acquaintances who were not explicitly endorsed were associated with greater N170 in controls suggesting that this was a pre-conscious mnemonic effect. Importantly, despite being posteriorly-distributed, this familiarity marker was reduced in patients with vmPFC lesions and the extent of the reduction predicted the severity of impairment in making rapid accurate memory decisions.

According to our schema hypotheses, instantiating aspects of a self-schema that makes accessible the relevant prior knowledge precedes stimulus presentation and guides perception and classification once the stimuli appear. Instantiating a corrupted or over-extended schema, as we hypothesize occurs in people with mPFC damage, leads to poor classification, or over-inclusion of items within the relevant category. To test this hypothesis, here we examined pre-stimulus frequency oscillations in the period before face presentation and related them to later performance on the personal familiarity task, as well as to modulations of the N170. We chose to look at frequency oscillations because lower band interregional coherence is functionally relevant for integration of information in distributed brain networks.

Altered schema may also affect post-stimulus monitoring or control processes in the personal familiarity task since responses are based on the output of the comparisons between schema-related knowledge and stimuli to determine if they exceed a given criterion, and if they do, to emit a response. To test for these post-stimulus effects, we examined the relation between the N170 and an anterior, positive modulation of the ERP that was also missing in patients with vmPFC damage. Similar modulations, which occur about 230 msec after stimulus onset (P230), had been implicated by Schnider, Valenza, Morand, and Michel (2002) in inhibitory control of task irrelevant information, a process they term *reality filtering* and propose to be mediated by medial posterior orbitofrontal cortex. If the P230 reflects the efficacy of the mPFC in monitoring or control processes, then its amplitude should be related to modulations of the N170 by personal familiarity. Last, if schema instantiation is a determining factor in post-stimulus monitoring, then there should be a significant relation between the P230 and pre-stimulus interregional coherence measures, which would track the integrity of the vmPFC.

2. Method

The development of new theories that emphasize the role of the vmPFC in schema processing, and in particular in schema instantiation that influences the way incoming information is processed, motivated us to analyze the pre-cue epoch of EEG. We previously described the methods to derive behavioral and post-cue ERP data, and reported the results of those analyses (Gilboa et al., 2009). Below are a brief overview and a more detailed description of the new analyses we performed.

2.1. *Participants*

Eight patients with lesions to the vmPFC following rupture of an ACoA aneurysm and eight healthy controls matched for age, gender and education were recruited from the Rotman Research Institute's subject pool. The maximum overlap of patients' lesions was in the ventral medial aspect of the prefrontal cortex, encroaching on basal forebrain in several patients (Fig. 1 in Gilboa et al., 2009). That said, patients varied in the exact location of lesion as well as its laterality. Whether lesions to specific locations within the mPFC or to a specific hemisphere tend to lead to greater deficits in behavior or in physiological abnormalities cannot be determined based on this group of patients alone.

Patients and controls were matched for age (51.12 and 54.50, respectively), years of education (15.75 and 16.50, respectively), sex (3 and 4 females, respectively), handedness (6 and 7 right handed, respectively), and estimated IQ based on performance on the Shipley Institute of Living Scale (103.25 and 106.37, respectively). Patients ranged with regard to the time since injury from four months to more than eight years post surgery with the average being around two years.

Patients had varying degrees of executive and anterograde memory impairments as determined by neuropsychological testing, but all were at least moderately impaired on one domain of memory (delayed verbal or delayed visual memory) and/or had executive dysfunction (Fig. 2 in Gilboa et al., 2009). More relevant to the task at hand is patients' autobiographical memory. Retrograde memory, as measured by the Autobiographical Memory Interview (AMI; Kopelman, Wilson, & Baddeley, 1990), was within normal range for both episodic (event) memory and personal semantic facts. By contrast, for anterograde (recent) autobiographical memory there was more variation across patients, with some of them showing impairments on both parts of the AMI.

As in our previous studies, we distinguish patients who spontaneously confabulate both from those who simply have high rates of erroneous responses to direct close-ended questioning and from those patients who have high rates of

intrusion on tests of recall or recognition. We incorporate case series methodologies in the present study to describe variations in performance and electrophysiology in relation to clinical symptomatology (Rosenbaum, Gilboa, & Moscovitch, 2014). In identifying spontaneous confabulators we rely on clinical reports and observations, and on elaborative confabulatory responses to open-ended questions that are similar to confabulations that would occur during naturalistic interactions. One patient who had an Anterior Communicating Artery (ACoA) aneurysm rupture 7 years earlier was still confabulating at the time of testing. Another patient (8 years post-surgery) was not actively confabulating as far as we know, but he spontaneously described himself as having false memories which he did not trust and which had led him to be very cautious about his memory and consult his friends and relatives regularly about events he thought he remembered. Within our theoretical framework, he is considered to have impaired automatic monitoring, for which he compensates by exerting meta-mnemonic conservative control process (Hebscher & Gilboa, 2016; Hebscher et al., 2015). In Figs. 3 and 7, these two patients are denoted by large black triangles. A third patient (18 months post-surgery) had a history of confabulation in the months immediately after his surgery, confirmed by interviews with family and clinical staff, but denied any current false memories. He is denoted by a small black triangle in Figs. 3 and 7. All procedures were approved by the Institutional Review Boards of Baycrest hospital and all participants signed an informed consent form prior to the beginning of the study.

2.2. Experimental design and materials

The experimental stimuli in the present study were comprised of three kinds of gray-scaled photos of faces:

1. *Personally familiar faces*: These were faces cropped out of pictures from family albums, covering the lifetime of the participant. Faces were cropped only from pictures in which the participant appeared so as to ensure that all of the persons represented in the picture were familiar. Still, familiarity of the faces varied, as some of the faces were of close relatives whereas others could be of old school teachers, classmates, work colleagues, etc. Moreover, faces of relatives could be taken from different periods of the subject's life (e.g., a childhood picture of a sibling). The same person's face could serve as a stimulus up to twice, provided that pictures from very different time periods were used (i.e., separated by at least 20 years). The number of stimuli to which participants were exposed varied somewhat among participants ($M = 133.56$, $SD = 16.82$, range = 108–156). There was no consistent group difference in the overall number of photos [$t(14) = -1.41$; $p > .05$] between controls ($M = 129.37$; $SD = 12.69$) and patients ($M = 137.75$; $SD = 11.01$).
2. *Famous faces*: There were 115 faces of famous personalities in naturalistic poses so as to match the faces from the other conditions. In all conditions faces could be in full frontal view, half views, profiles etc. and faces could express emotion, reflect a speech act, etc. Additionally, we purposely selected a portion of the pictures of famous faces which were of low quality (blurry, dark etc.) so as to mimic the pictures of faces from family members and acquaintance, which were sometimes taken out of very old or low quality pictures.
3. *Unfamiliar faces*: Participants also viewed 120 faces of unfamiliar people, who were selected out of the personal stimuli of other participants. There were faces of people of different chronological ages and from different decades so as to match the distribution of personally familiar and famous faces.

2.3. Task and procedures

Participants were told that they would view faces of people from their personal pasts, faces of famous personalities and faces of unfamiliar people. They were told that they should respond with 'yes' (left mouse-key; pointing finger) only for people they had personally encountered and with 'no' (right mouse-key; middle finger) for all others, regardless of whether they know who they are or not. Each trial was preceded by a fixation with the question "have you personally met this individual before?" and mapping of the response keys at the bottom, which appeared for 1000 msec. The question and mapping of the keys remained on screen the whole time. Following fixation, the faces appeared for 600 msec and after the face disappeared, participants had an additional time window of 2000 msec to respond, during which only the response key mapping were on screen. On average, participants responded between 750 and 1200 msec post-stimulus onset and the empty screen remained until the onset of the next fixation that signaled the beginning of a new trial. Participants were encouraged to respond as quickly as possible but not at the expense of accuracy.

Following the application of the electrode-cap, the experiment was run within a single session lasting about 10–12 min. A pause was included mid-way through the session (after approximately 140 stimuli, depending on the number of personally familiar faces). During the break the experimenter entered the chamber to speak to the participants, to ensure they were not too fatigued and to encourage them. Participants in fact reported enjoying the experiment, and in particular they reported enjoying seeing faces of persons they hadn't thought about in a long time.

2.4. ERP recordings and analysis

Continuous EEG was recorded with Neuroscan software using a Synamps neural amplifier (Compumedics, El Paso, TX, USA) and 64-channel tin electrode caps (Electro-Cap International), with electrodes placed according to the 10–20 system (Jasper, 1958). EEG recordings were made at a sample rate of 500 Hz, using a Cz reference. The continuous EEG recordings were filtered on line between .05 and 100 Hz. During the recording, electrodes placed at the outer canthi and the superior and inferior orbit monitored vertical and horizontal eye movements.

Analyses were performed using Brain Electrical Source Analysis software (BESA Research 6.1 and BESA statistics 2.0, MEGIS software GmbH, Gräfelfing, Germany). The continuous

EEG files were visually inspected for channels displaying faulty recordings and these were either interpolated or ignored (if they were around the rim of the cap) and re-referenced to a common-average reference. Eye movement artifacts were corrected using horizontal and vertical electrooculography (EOG) measurements. Artifact correction was applied when EOG signal exceeded the thresholds of 150 μV (horizontal EOG) or 250 μV (vertical EOG). Trials were sorted by stimulus type: correct rejection of famous and non-famous faces and hits of personal acquaintances; we ignored misses and false alarms in the present investigation. Accordingly the continuous EEG was segmented into 2100 msec epochs including a 1500 msec pre-stimulus window (500 msec empty screen and 1000 msec fixation) and the 600 msec of stimulus visual presentation. Artifact rejection of individual trials was performed for trials that contained maximum amplitudes higher than 120 μV, maximum amplitude difference between two neighboring samples was more than 75 μV or maximum signal amplitude lower than .01 μV. Artifact free signals were high pass filtered with .2 Hz zero-phase filter (6 dB/octave). This was conducted automatically with channel/trial exclusions checked manually for each file before processing could continue. Across healthy participants, the number of included trials for Hits, correct rejections (CRs) famous and CRs non-famous varied between 68 and 122 for each category. For the patients, the number of included trials for Hits, CRs famous and CRs non-famous varied between 36 and 116 for each category.

For the time-frequency analysis, single trial data corresponding to each condition were transformed into the time-frequency domain examining the changes in power, relative to the baseline period. Because of our hypothesis that mPFC would interact with posterior neocortical structures through lower band frequencies, and in particular the theta (4 Hz–8 Hz) and beta (14 Hz–30 Hz) frequency bands, we focused our analyses on frequencies between 2 Hz and 30 Hz (in 2 Hz increments), for each trial. To examine pre-stimulus oscillatory coherence we defined the 500 msec empty screen before the onset of fixation as baseline and analyzed the 500 msec fixation immediately prior to face onset. The power changes were then averaged across trials, and average ERP signal was subtracted before further analyses. The time-frequency data of each participant were then analyzed by source in BESA using the iterative 3D source imaging method CLARA (Classical LORETA Recursively Applied) separately for theta and beta. The CLARA approach iteratively localizes activity to the constrained regions identified from the previous solution. Three iterations were computed using the default voxel dimension of 7 mm^3 and 1% regularization constant. The solution was computed using an adult realistic head model in BESA 6.1 and registered against the standardized BESA finite element model, which was created from the average of 24 individual anatomical magnetic resonance imagings (MRIs) in the Talairach-Tournoux coordinate space. Group differences in surface-level time-frequency and in source solution of frequency bands were tested using *t*-test implemented in BESA statistics 2.0. A preliminary Student's *t* test between conditions per data point was calculated followed by 1000 parameter-free permutation testing in combination with data clustering to correct for multiple comparisons.

To examine cross-regional coherence we used Dynamic Imaging of Coherent Sources (DICS; Gross et al., 2001) for the theta time-frequency range (4–8 Hz) which significantly differed across groups in the time-frequency analysis. We used the coordinate in mPFC derived from the source analysis. Because patients had lesions to parts of the mPFC, we also performed source coherence analysis for a coordinate in right posterior inferotemporal cortex in order to probe the possibility that posterior neural populations display synchronized/desynchronized firing patterns with other, non-mPFC, cortical areas. The coordinate used for this analysis was derived from a source analysis of the N170. Statistical analyses of group differences in source coherence were performed in the same way as the time-frequency source analyses.

3. Results

3.1. Behavioral data

A detailed report of the behavioral results was previously published (Gilboa et al., 2009). Overall, both patients and controls were able to distinguish faces of persons from their own past and other faces, although patients were moderately impaired compared to controls, with d' for patients ranging from ~1.50 to 3.50 and controls' from ~3 to 4.50 (Fig. 7C). There was a significant within-subject main effect of accuracy, with no accuracy by group interaction because hit rates were significantly lower than correct rejections of famous and non-famous faces. These lower hit rates reflected the fact that face pictures were cropped out of pictures from participants' entire lives and often involved faces from many years ago (e.g., class mates, work colleagues etc.). Patients and controls had similar reaction times (RTs), and both groups could reject famous and unfamiliar faces more quickly than they could endorse personally familiar faces.

3.2. ERP results

Detailed analyses of the face-related ERP's appear in the previous report. Here, in addition to the detailed pre-cue time-frequency analyses, we added a source analysis for the N170 component using CLARA and a correlational analysis of the N170 familiarity index and the early anterior positive modulation at 230–260 msec.

Fig. 2 shows the result of the source analysis for the N170 in right inferotemporal cortex. The maximum coordinate of the source solution is ($x = 48$, $y = -30$, $z = -20$) and the extent of the source solution overlaps with the region typically reported for the fusiform face area in neuroimaging studies (max: $x = 42$ $y = -50$ $z = -22$; meta-analysis of 116 studies using the term "fusiform face" on neurosynth.org). The inferotemporal cortex is important for representational aspects of face identity, particularly in more anterior portions. We previously reported N170 sensitivity to face familiarity in controls, regardless of explicit recognition on this task. Together, these results suggest an early, automatic, post-stimulus marker of prior knowledge originating from posterior neocortical regions. Importantly, that posterior familiarity marker is missing in vmPFC patients even though their damage is

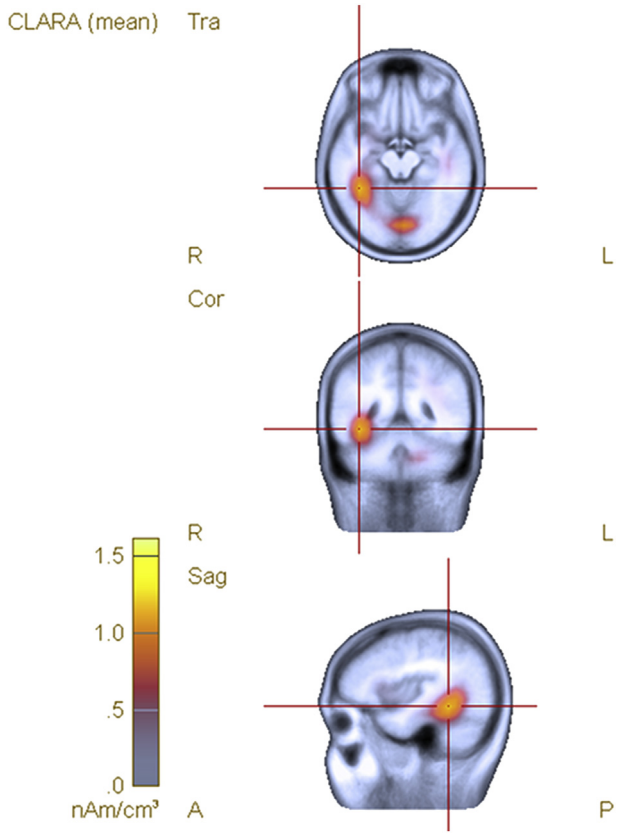

Figure 2 – Source solution using CLARA for the N170 ERP component averaged over the time window of 160–180 msec.

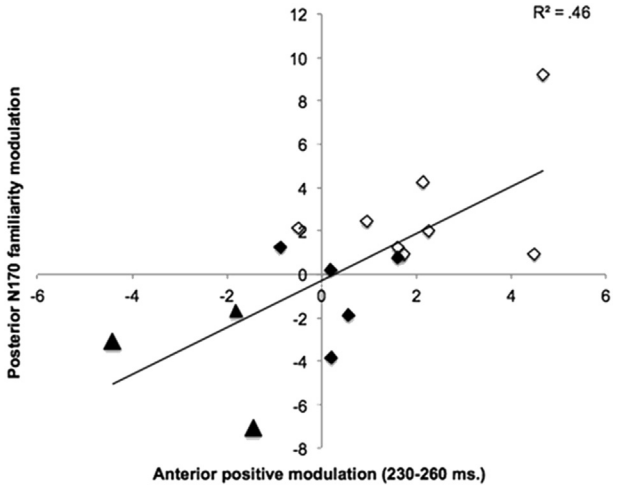

Figure 3 – Correlations between N170 (CB2, P8) familiarity index (acquaintances *vs* unfamiliar) and positive waveform modulation over anterior electrodes (FP1, FPZ, FP2). Open diamonds are healthy controls, black diamonds are vmPFC patients and black triangles are vmPFC patients with histories of confabulation.

anterior. Early posterior activity that depends on anterior structures could reflect either very rapid top-down modulation of posterior cortex by mPFC (Bar et al., 2006) or the effect of pre-cue biases that lead to privileged processing of context-relevant stimuli (Uretzky & Gilboa, 2010; see below), or both effects.

Our monitoring hypothesis suggests that the automatic FOR arises from a fit between activated representations from long-term memory and a memory cue, and that the better the fit, the stronger the FOR (Fig. 1B). We hypothesized that an early anterior positive modulation that differed between patients and controls might be a neurophysiological marker of an early monitoring process (Gilboa et al., 2009). That it was correlated with RTs for accurate memory decisions supported that hypothesis, but whether it was also related to posterior representations was not reported. Fig. 3 presents the correlation between the N170 familiarity index (difference in amplitude between the N170 for known and unknown persons) and the P230 anterior positive modulation. Larger differences in N170 were associated with larger anterior positivity, consistent with our predictions. Moreover, patients with confabulation tended to have the most negative anterior modulation and showed a reversed familiarity index. These two components are relatively close in time (~60 msec) and, therefore, may not represent distinct neural events but instead may reflect biomechanic relationships or signals from the same neural generator. We, therefore, also correlated the condition-specific N170 amplitudes (correct rejections of famous and non-famous persons) with the P230 amplitude. There was no relationship between the P230 and the N170 for famous faces ($r = .14$, $p > .1$) but a strong correlation between the one associated with non-famous faces and the P230 ($r = .6$, $p < .01$). There was a significant difference between the two correlations [Williams-Steiger *t*-test: $t(13) = -3.32$, $p < .05$]. This suggests that the association between the posterior N170 familiarity index and anterior P230 is related to specific cognitive indices of the ERP rather than to general characteristics of the signal.

3.3. *Pre-stimulus time-frequency analysis*

A major focus of the present analysis pertains to the idea that the mPFC biases posterior representations in a context-sensitive manner such that relevant schemas are preferentially activated to enable more efficient processing of information (Fig. 1A). We predicted this schema effect would be reflected in changes in mPFC-posterior neocortex coherence in low frequency bands prior to stimulus onset. In line with our hypothesis, there was a significant difference between controls and patients in relative power changes averaged across all electrodes ($p = .04$, permutation test) that was apparent in the low frequency bands (4–16 Hz). Moreover, controls demonstrated decreased power in that cluster (controls = −.06; patients = .01) in the 500 msec leading to stimulus onset compared to pre-fixation baseline (Fig. 4A). Source analysis of theta (4–8), alpha (9–13) and beta (14–30) power desynchronization revealed no significant clusters for the alpha and beta power band, but significant medial prefrontal theta desynchronization in controls but not in patients (Fig. 4B). We, therefore, focused our interregional coherence analyses on the theta power band.

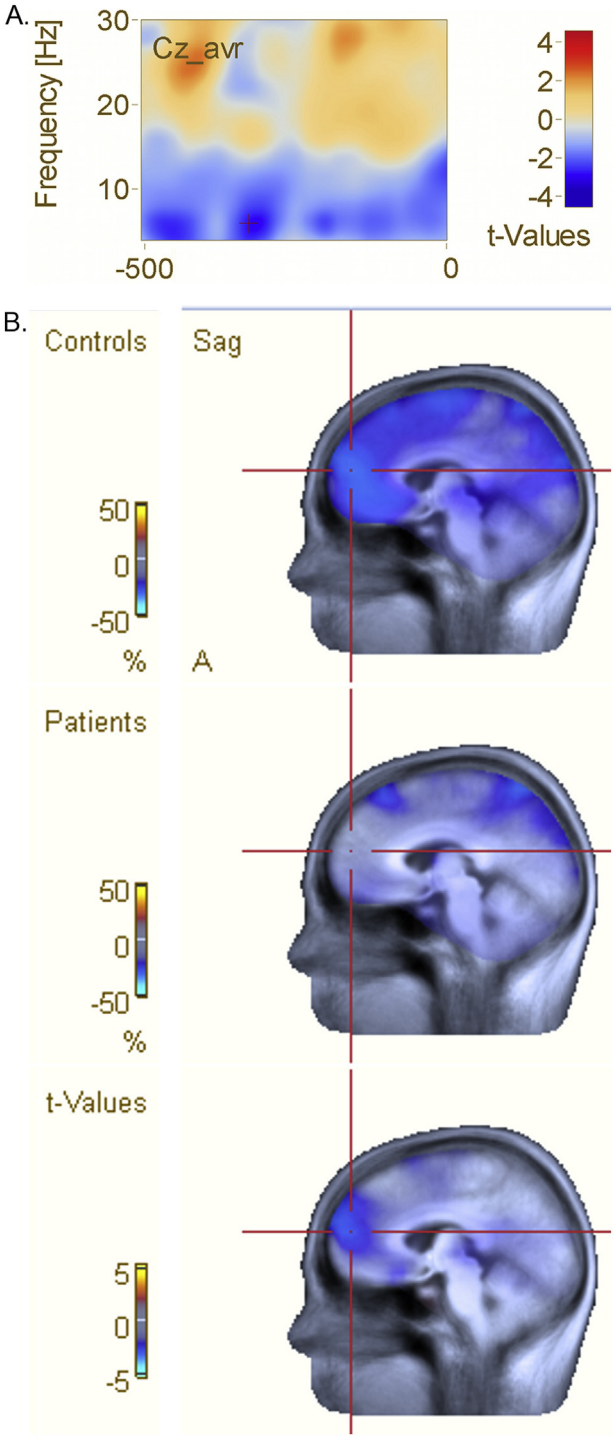

Figure 4 – (A) Significant time–frequency differences in relative power changes averaged over all electrodes between the healthy control group and vmPFC patients. (B) Source solution for theta band (4–8 Hz) power averaged over the pre-cue epoch (−500 to 0 msec) showing controls, patients and the statistical t-map of the differences between them.

For our main source coherence analysis we used the coordinates in mPFC identified as the peak difference between groups ($x = 5$ $y = 54$ $z = 16$). A large significant cluster of desynchronized cross-regional theta ($p = .01$, permutation test) distinguished controls ($M = -20.95$; Fig. 5A) from patients ($M = -5.29$; Fig. 5B). The cluster had peak t-values in bilateral mPFC and bilateral inferotemporal and lateral temporal cortices, more pronounced on the left (Fig. 5C). These data are consistent with the idea of cross-regional interactions between mPFC and posterior and lateral temporal cortices that are typically considered to support long-term memory representations.

Our primary analysis was based on an unbiased whole-brain analysis that pointed to mPFC as the generator of theta desynchronization during the pre-cue period. We were concerned that group differences in mPFC cross-regional coherence patterns could be merely a reflection of atrophied neural tissue in that region in the patient group. We, therefore, also investigated whether posterior regions demonstrated synchrony or desynchrony with other cortical or subcortical structures. Patients did not have posterior lesions, and presumably could form representational preparedness through compensatory processes. To that end we used the posterior inferotemporal coordinate that source analyses suggested was associated with post-stimulus facial recognition N170 ($x = 48$, $y = -30$, $z = -20$; Fig. 2). We reasoned that the pre-cue network activity, presumably reflecting biased representation of relevant prior knowledge, would be important for generating that mnemonic bias. Moreover, this is a part of the posterior cortical network that processes facial identity and so would be highly relevant for our task. Using this posterior coordinate again revealed large areas of inter-regional desynchrony in the theta frequency band that differed between controls and patients ($p = .009$, permutation test) with controls again showing more desynchronization ($M = -21.06$; Fig. 6A) than patients ($M = -5.41$; Fig. 6B). The distribution of the differences was similar to the one identified for the mPFC seed involving bilateral inferotemporal and lateral temporal, more on the left as well, and medial prefrontal white and gray matter bilaterally (Fig. 6C).

Last, to investigate whether this pre-stimulus desynchronization was associated with stimulus processing at the physiological and behavioral levels, we performed correlational analyses between individual cluster-level desynchronization scores and the N170 familiarity-related modulation, the P230 anterior modulation, as well as with accuracy (d') and RTs. Consistent with our hypothesis there were significant correlations between pre-stimulus desynchronization and the posterior N170 familiarity index ($r = -.6$, $p = .01$; Fig. 7A), the P230 ($r = -.62$, $p = .01$; Fig. 7B) and with accuracy ($r = -.58$, $p = .02$; Fig. 7C), but we did not find a correlation with RTs ($r = .37$, $p = .16$).

3.4. *Control analyses*

(i) Our patient group (and by extension the matched controls) varied on several dimensions that could contribute to the patterns of behavior and physiological differences beyond the lesions themselves. These include age, known to affect prefrontal structure and function, as well as education and intelligence that could influence familiarity with famous faces or some executive functions. To test these possibilities we correlated age, education and estimated IQ with

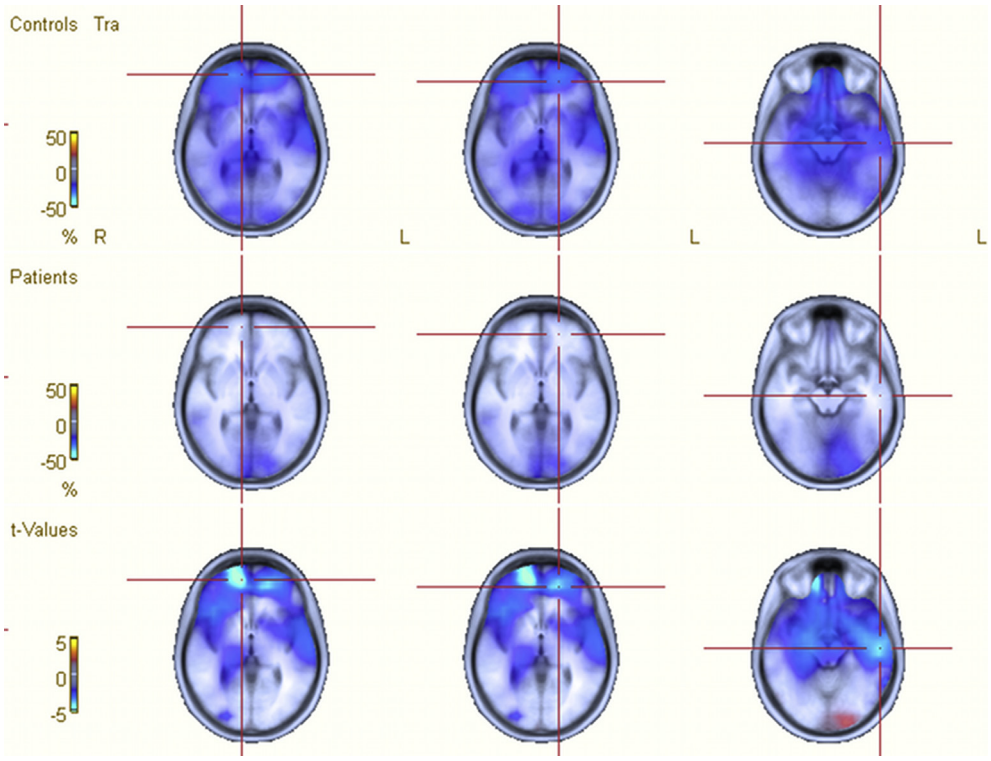

Figure 5 – Medial prefrontal DICS for theta coherence desynchronization in (A) controls and (B) patients, as well as (C) significant clusters of difference in medial prefrontal and inferior and lateral temporal cortices.

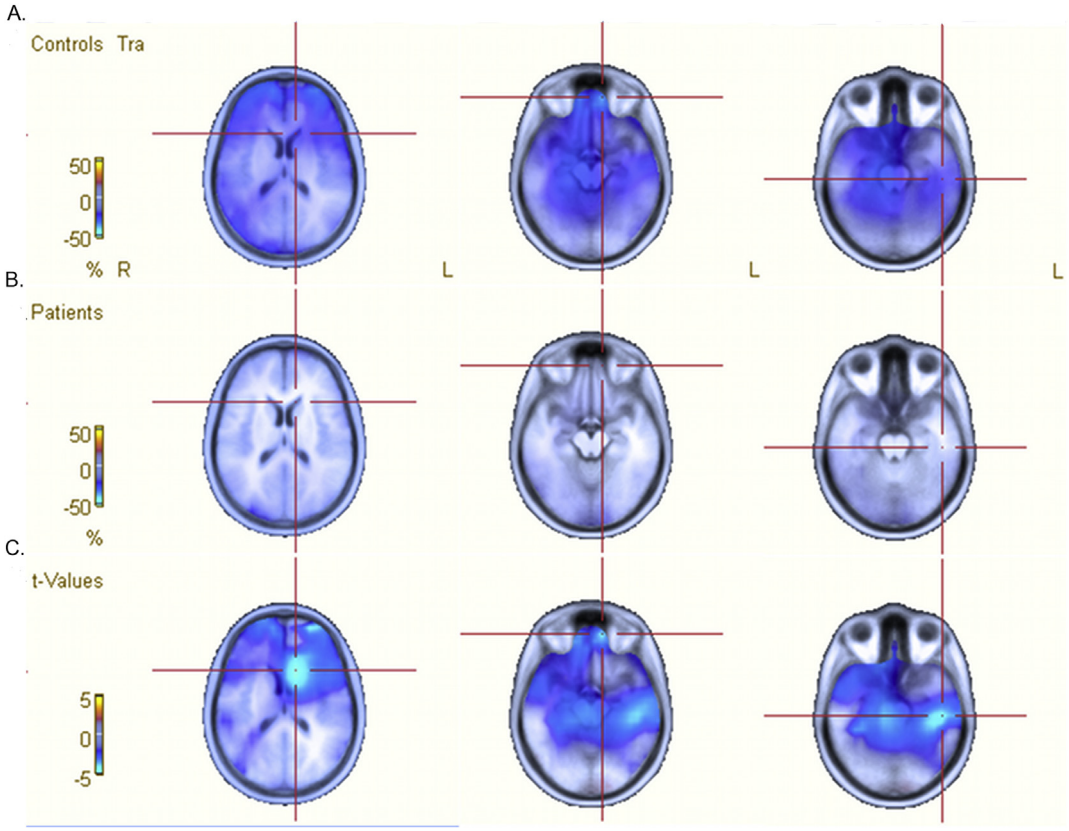

Figure 6 – Right posterior inferotemporal seed DICS for theta coherence desynchronization in controls (A) and patients (B), as well as cluster-corrected peaks of significant differences (C) demonstrating differences in subcortical, vmPFC and left lateral temporal cortices.

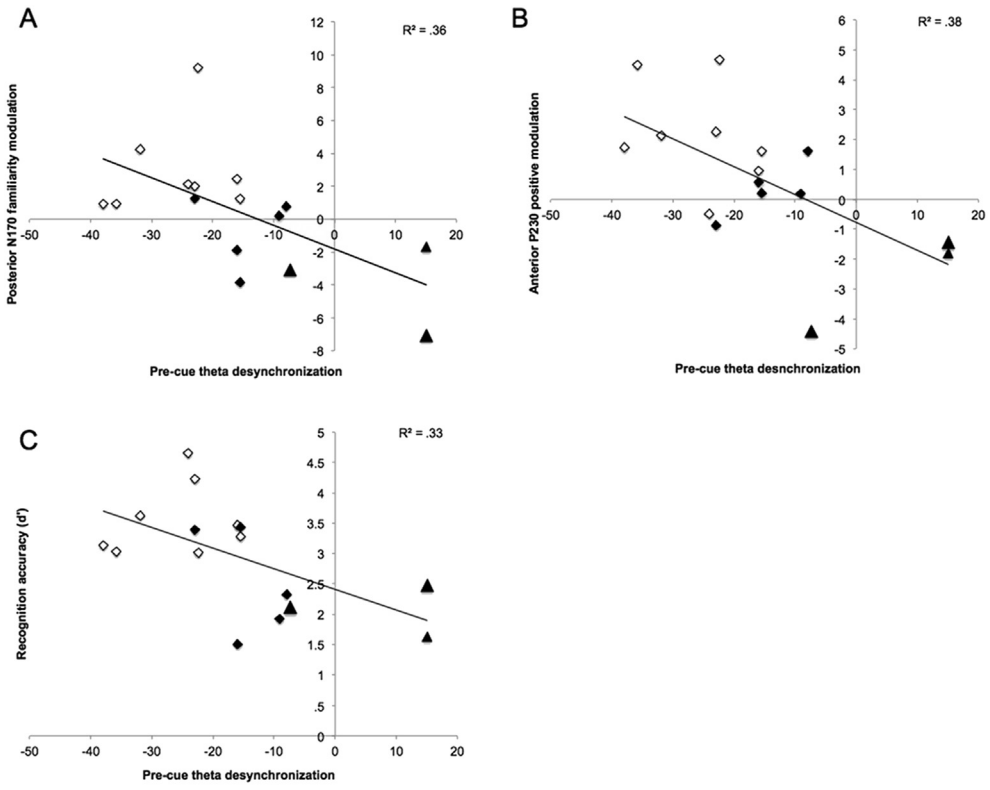

Figure 7 – Correlations of medial prefrontal pre-cue theta desynchronization with (A) familiarity-related modulation of N170 (acquaintances *vs* non-famous) (B) anterior P230 modulation and (C) accuracy (d' acquaintances *vs* famous). Open diamonds are healthy controls, black diamonds are vmPFC patients and black triangles are vmPFC patients with histories of confabulation.

behavioral (d' recognition sensitivity and RT's) and major electrophysiological components (pre-cue theta desynchronization, N170 familiarity index and P230). Age and education did not significantly correlate with any of the behavioral or electrophysiological measures. IQ significantly correlated with pre-cue theta desynchronization ($r = -.58$, $p < .05$) and marginally correlated with the N170 familiarity index ($r = .42$, $p = .1$). While speculative, such correlations could reflect known relationships between IQ and efficient use of knowledge schemas during intuitive and insight-based problem solving (McCrae, 2010).

(ii) Unlike controls, patients' skulls may contain deformities due to surgical procedures and such bone defects could potentially influence whole-scalp EEG analyses as well as source solutions for specific components. To test the possible contribution of such biomechanic influences, we compared the N170 for both groups using all conditions (i.e., regardless of face type and response accuracy). The reason we chose the N170 and the entire range of stimuli was to get the most reliable ERP component and best SNR to maximize our chances of finding group differences if they exist. Note that although there was a group by condition interaction for the N170 (familiarity index in controls only), there was no group main effect (Gilboa et al., 2009). Fig. 8 presents the scalp density maps for patients and controls as well as the source solution for the N170 separately for patients and controls. As can be seen, both scalp level maps and source localization using CLARA are very similar between groups, even for one of the most reliable ERP components suggesting that if bone defects affected the topography and source solution of the EEG, that effect is minimal.

(iii). As is evident from the scatter plots in Figs. 3 and 7, patients with histories of confabulation tended to show more extreme difference in electrophysiological patterns compared to healthy controls and possibly non-confabulating patients. We formally investigated this impression by testing whether patients with confabulation statistically differ in the extent to which they display (1) pre-cue theta desynchronization (2) post-cue N170 familiarity index (3) P230. Non-parametric Kruskal–Wallis tests demonstrated significant differences across the three groups (controls, confabulators, non-confabulators) on all three components [$\chi^2(2) = 9.94$, $p < .01$; $\chi^2(2) = 10.02$, $p < .01$; χ^2 (2) $= 9.69$, $p < .01$, respectively]. Specific contrasts revealed that this difference occurred because non-confabulating patients had less desynchronization, and smaller P230 than controls (Mann–Whitney $U = 0$, $Z = -2.36$ $p < .05$ for both tests). Further, confabulating patients had less desynchronization and smaller P230 than both controls and non-confabulating patients ($U = 0$, $Z = -2.45$, $p < .05$ and $U = 0$, $Z = -2.23$, $p < .05$, respectively for both components), and smaller N170 familiarity index than

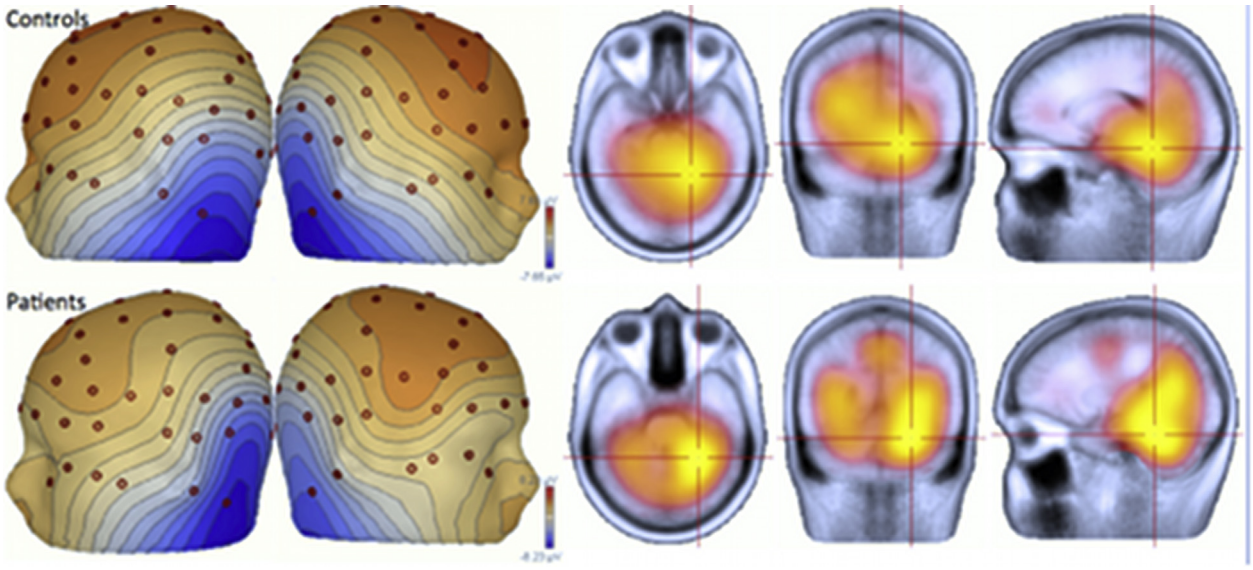

Figure 8 – Scalp density distributions (left) and source solutions using CLARA (right) for the N170 across all conditions for controls (top) and patients (bottom).

controls ($U = 0$, $Z = -2.45$, $p < .05$). Thus, it appears that the presence of confabulation is associated with more extreme pre-cue and post-cue neurophysiological markers of prior knowledge processing.

4. Discussion

Our strategic retrieval framework of confabulation and retrieval-related functions of the mPFC has identified two potential roles for prior knowledge, or schema memory. The first is pre-stimulus schema instantiation in which the mPFC helps maintain a context-relevant schema that influences processing of ongoing environmental stimuli or self-generated cues that can guide retrieval (Fig. 1A). The second is post-stimulus, automatic monitoring (FOR) which depends in part on the correspondence between retrieved memories and activated schemas in posterior neocortical regions (Fig. 1B). Analyzing ERPs and time-frequency oscillations, and relating them to each other and behavior, we found evidence consistent with both schema functions. In addition, meta-mnemonic control mechanisms, which have not been investigated here, operate on the output of these processes to determine whether to act on the basis of the retrieved information (Fig. 1C,D).

4.1. *Pre-stimulus schema instantiation*

Schema instantiation is the process of activating and sustaining a general 'template' representing a knowledge structure and biasing information processing from the input stream to be consistent with activated schema variables (Ghosh & Gilboa, 2014; Thorndyke & Yekovich, 1980). Neurophysiologically, it involves contextually sensitive biasing of posterior representations. This process, which is mediated by mPFC, should be protracted and appear as tonic effects in the time-frequency domain. Consistent with this view, we found that controls, but not patients with vmPFC lesions, demonstrated changes in low-frequency power and mPFC interregional coherence in the period preceding the stimulus, mostly in the theta frequency range. These changes reflected desynchronized activity across cortical regions necessary for making personal familiarity judgments (Collins & Olson, 2014). These regions include the bilateral inferotemporal cortex which is implicated in perceptual face processing (Rossion & Gauthier, 2002) and left lateral temporal cortex which is associated with corresponding semantic processing (Binder, Desai, Graves, & Conant, 2009; Collins & Olson, 2014).

Ample evidence accumulated in recent years has shown that both synchronous and desynchronous neuronal oscillations are critical for efficient cognitive and behavioral processing (Hanslmayr, Staresina, & Bowman, 2016; Klimesch, Sauseng, Hanslmayr, Gruber, & Freunberger, 2007). Our data attest to the functional importance of active pre-stimulus desynchronization by showing that it correlated with both the familiarity index N170 modulation and with response accuracy in the personal familiarity task. We interpret these results as reflecting the neurophysiological mechanism by which mPFC instantiates and maintains schemas relevant for processing upcoming stimuli (see Ghosh et al., 2014). In the memory domain, it has been proposed that such pre-stimulus activity reflects the activation of context-sensitive information by attentional mechanisms (Cohen et al., 2015), with the relevant context here being those aspects of the self-schema that can be applied to the task at hand. Consistent with this view, patients' smaller desynchronization might be indicative of either less efficient suppression of distractor information or activation of irrelevant information, which can lead to over-extension of schemas.

Alternatively, smaller desynchronization may reflect less efficient or slower transitions between processing states or schemas, thereby leading to erroneous acceptance of stimuli that were relevant to a previously activated state but not the current one (cf. Schnider, 2008). Patients with confabulation or

a history of confabulation in the present study tended to display the least pre-stimulus desynchronization (and even theta synchrony), reversed familiarity related N170 amplitudes and negative, rather than positive P230. These findings are consistent with the proposal that confabulation results from the intrusion of "input templates" – powerful generic memories that serve as "starting values" for the recollection process (Burgess & McNeil, 1999; Burgess & Shallice, 1996). Failure to "deactivate" input templates when they are inappropriate could be mediated by the absence of active desynchronization. A similar idea, using different mechanisms, has been proposed by Shallice and Cooper (2012).

It has been suggested that decreased interregional coherence can support better expression of relevant neural codes that can then be processed more efficiently downstream (e.g., by the hippocampus) during memory encoding and retrieval (Hanslmayr et al., 2016).

The neurophysiological mechanisms by which desynchronization can lead to more efficient neural processing are unclear, although three candidate mechanisms (Hanslmayr et al., 2016) might be relevant to our data and model. One idea is that desynchronization, in particular of low frequencies, could serve as a gating mechanism that enables increased firing rates during stimulus processing. A second possibility is that decreased low frequency synchrony leads to decreased variance in neural firing and consequently increases the reliability of neural code. Finally, decreased power and coherence decorrelates ongoing neural activity and allows stimulus-specific flexible phase adjustment. Whatever the mechanism, mPFC could bias processing of stimuli in a schema-consistent manner by maintaining tonic low-frequency desynchronization over context-relevant posterior neural networks. Indeed, the tonic pre-stimulus desynchronizations we observed appears to influence the way stimuli are processed very early post-stimulus and are associated with context-sensitive accurate responses in our personal familiarity task.

4.2. *Post-stimulus effects*

Schema instantiation mediated by the mPFC may also contribute to post-stimulus monitoring which requires comparison of the stimulus with the schema (Gilboa, 2004, 2010; Gilboa et al., 2006). We have previously argued that mPFC lesions could lead to activation of corrupted or over-extended schema (Ghosh & Gilboa, 2014; Ghosh et al., 2014) and consequently, indiscriminate, non-informative FOR and confidence signal (Hebscher & Gilboa, 2016; Hebscher et al., 2015). FOR emerges as a result of strong correspondences between retrieval cues and activated long-term memory traces; more extensive long-term memory associative networks give rise to stronger FOR. The task in this study required activation of a schema related to personal experiences with people, which may overlap with a self-schema and, therefore, be a potent catalyst of FOR. This interpretation could explain why confabulations are most often observed in the autobiographical domain, and why these are associated with the strongest conviction and confidence in their veracity (Gilboa & Moscovitch, 2002; Moscovitch, 1989), even when cues are highly implausible and presented via recognition (Gilboa et al., 2006); cf. (Kan et al., 2010).

Our present analyses are consistent with the idea that activated schemas affect early monitoring processes by showing a significant correlation between the N170 familiarity index and the later frontally distributed P230. In our previous study (Gilboa et al., 2009), we noted that this early frontal component was correlated with RTs for accurate responses (but not overall accuracy), consistent with the idea that it is associated with a rapid form of veracity monitoring. Importantly, here we show that the P230 is also correlated with pre-stimulus coherence desynchronization, consistent with the idea that the earlier instantiated schemas influence subsequent monitoring process. Patients with vmPFC damage and, in particular, those with confabulation appear to be missing the P230 phasic modulation of the ERP. Instead, they display a prolonged late anterior negativity that correlates with accuracy and could reflect later, more controlled, compensatory processes.

Confabulators' difficulty in maintaining a coherent schema structure or in inhibiting irrelevant schemas from intruding on their responses could be partially an outcome of a tendency to process task irrelevant information (Ciaramelli, Ghetti, & Borsotti, 2009). Ciaramelli et al. found that dividing attention during memory retrieval decreased confabulators' false memories. Healthy controls demonstrated the expected reverse pattern. These findings are consistent with the hypothesis that obstructing confabulators' excessive processing of irrelevant information through divided attention should lead to improved performance. Excessive processing of extraneous stimuli (external or internal) could lead to activation of associated information which, in turn, could contribute to inappropriate FOR. Desynchronization of posterior cortical areas could be a mechanism that inhibits processing of irrelevant information, allowing the maintenance of uninterrupted activation of relevant long-term memory representations.

These results are not incompatible with a view of confabulation advanced by Schnider which he termed *impaired orbitofrontal reality filtering* (Schnider, 2003, 2008; Schnider et al., 2002) which depends on intact extinction processes (Nahum, Ptak, Leemann, & Schnider, 2009, 2013). According to Schnider, the posterior orbitofrontal cortex is needed to restrict processing to currently relevant information and inhibit processing of task irrelevant information, thereby anchoring the individual in the current reality. Accordingly, failure to extinguish previously rewarded, but currently irrelevant, stimuli is considered the basis of behavioral confabulation and disorientation. As a result, damage to the orbitofrontal cortex allows irrelevant information to be processed, leading not only to a distortion of reality evident in the confabulator's remarks, but also to action based on these erroneous beliefs. In our model, such effects are ascribed to corrupted vmPFC-mediated schema representations, which operate at both encoding and retrieval, as well as to posterior OFC-mediated defective control processes at retrieval (Hebscher & Gilboa, 2016; Hebscher et al., 2015). Interestingly, according to Schnider, posterior OFC may function by rapid, phasic, desynchronization of post-stimulus coordinated activity in networks that represent currently irrelevant thoughts (Schnider, 2003, 2008; Schnider et al., 2002).

4.3. *Confabulation in other patient populations*

Our paper focused on the role of the vmPFC in normal and impaired schema memory, with confabulation serving as a case in point for failed schema processes. We end by considering a broader view of confabulation as a syndrome that appears in a variety of disorders.

Although confabulation is often associated with damage to the vmPFC and posterior OFC (Gilboa & Moscovitch, 2002; Schnider, 2003; Turner, Cipolotti, Yousry, & Shallice, 2008), there are many cases of confabulation in which such damage cannot be substantiated. For example, confabulation has been reported in patients with Alzheimer's disease, diencephalic lesions, multiple sclerosis and most commonly Wernicke-Korsakoff syndrome (WKS) (Gilboa & Moscovitch, 2002; Gilboa & Verfaellie, 2010). It is possible, as is the case with WKS patients, that the disorder affects the operations of vmPFC and damage can be detected with sensitive neuroimaging techniques and post-mortem (Kopelman, 2015). The same may be true of other patient populations, but we are not aware of research that has examined the possibility. Interestingly, dementia tends to lead to confabulation more prominently when vmPFC/OFC are affected early as in frontotemporal lobar degeneration or dementia (FTLD) (e.g., Mendez, Fras, Kremen, & Tsai, 2011) offering some support for this possibility. Moreover, our current investigation emphasizes the importance of a network approach to understanding confabulation (Fig. 1), suggesting that it may arise not just from direct damage to the vmPFC, but from damage to other regions of the core memory network or from poor coordination among them. The vmPFC plays a central role within this network, acting as a hub that coordinates and supports key functions in the multi-factorial mnemonic processes whose breakdown can lead to confabulation.

5. Conclusion

To answer the question with which we began – namely when errors occur, what prevents memory from being wildly inaccurate? – we proposed the *schema instantiation hypothesis* which emphasizes the critical role of strategic memory processes. To be more specific, instantiating a task or context relevant schema restricts memory search and responses to a domain that is consistent with the appropriate schema. We suggest that in addition to its phasic role after the stimulus has appeared, the mPFC may be involved prior to stimulus presentation in supporting protracted, privileged access of contextually relevant long-term representations to conscious processing. Damage to the mPFC disrupts these schema related processes both before and after stimulus presentation or memory recovery, leading to confabulation.

Acknowledgment

This research was conducted with support from the Canadian Foundation for Innovation/Ontario Innovation Trust Grant 1226 and the Heart and Stroke Foundation Centre for Stroke Recovery. A.G. is supported by Natural Sciences and Engineering Research Council of Canada (grant 405649) and M.M. is supported by grants from the Canadian Institutes for Health Research (grant MGP 6694) and from the Natural Sciences and Engineering Research Council of Canada (grant A 8347).

REFERENCES

Attali, E., De Anna, F., Dubois, B., & Dalla Barba, G. (2009). Confabulation in Alzheimer's disease: Poor encoding and retrieval of over-learned information. *Brain, 132*(Pt 1), 204–212.

Bajo, A., Fleminger, S., Metcalfe, C., & Kopelman, M. D. (2017). Confabulation: What is associated with its rise and fall? A study in brain injury. *Cortex, 87*, 31–43.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America, 103*(2), 449–454.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767–2796.

Brod, G., Lindenberger, U., Werkle-Bergner, M., & Shing, Y. L. (2015). Differences in the neural signature of remembering schema-congruent and schema-incongruent events. *Neuroimage, 117*, 358–366.

Burgess, P. W., & McNeil, J. E. (1999). Content-specific confabulation. *Cortex, 35*(2), 163–182.

Burgess, P. W., & Shallice, T. (1996). Confabulation and the control of recollection. *Memory, 4*(4), 359–411.

Ciaramelli, E., Ghetti, S., & Borsotti, M. (2009). Divided attention during retrieval suppresses false recognition in confabulation. *Cortex, 45*(2), 141–153.

Cohen, N., Pell, L., Edelson, M. G., Ben-Yakov, A., Pine, A., & Dudai, Y. (2015). Peri-encoding predictors of memory encoding and consolidation. *Neuroscience and Biobehavioral Reviews, 50*, 128–142.

Collins, J. A., & Olson, I. R. (2014). Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia, 61*, 65–79.

Craik, F., I.M., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., et al. (1999). In search of the self, a positron emission tomography study. *Psychological Science, 10*(1), 26–34.

Crutch, S. J., & Warrington, E. K. (2011). Taxonomic and thematic organisation of proper name conceptual knowledge. *Behavioural Neurology, 24*(4), 265–276.

Dalla Barba, G., Brazzarola, M., Marangoni, S., Barbera, C., & Zannoni, I. (2017). A longitudinal study of confabulation. *Cortex, 87*, 44–51.

Dalla Barba, G., & La Corte, V. (2013). The hippocampus, a time machine that makes errors. *Trends in Cognitive Sciences, 17*(3), 102–104.

Fotopoulou, A. (2010). The affective neuropsychology of confabulation and delusion. *Cognitive Neuropsychiatry, 15*(1), 38–63.

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia, 53*, 104–114.

Ghosh, V. E., Moscovitch, M., Melo Colella, B., & Gilboa, A. (2014). Schema representation in patients with ventromedial PFC lesions. *The Journal of Neuroscience, 34*(36), 12057–12070. http://dx.doi.org/10.1523/JNEUROSCI.0740-14.2014.

Gilboa, A. (2004). Autobiographical and episodic memory–one and the same? Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia, 42*(10), 1336–1349.

Gilboa, A. (2010). Strategic retrieval, confabulations, and delusions: Theory and data. *Cognitive Neuropsychiatry, 15*(1), 145–180.

Gilboa, A., Alain, C., He, Y., Stuss, D. T., & Moscovitch, M. (2009). Ventromedial prefrontal cortex lesions produce early functional alterations during remote memory retrieval. *The Journal of Neuroscience, 29*(15), 4871–4881.
Gilboa, A., Alain, C., Stuss, D. T., Melo, B., Miller, S., & Moscovitch, M. (2006). Mechanisms of spontaneous confabulations: A strategic retrieval account. *Brain, 129*(Pt 6), 1399–1414.
Gilboa, A., & Moscovitch, M. (2002). The cognitive neuroscience of confabulation: A review and a model. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), *Handbook of memory disorders* (2nd ed., pp. 315–342). London: John Wiley & Sons.
Gilboa, A., & Verfaellie, M. (2010). Telling it like it isn't: The cognitive neuroscience of confabulation. *Journal of the International Neuropsychological Society: JINS, 16*(6), 961–966.
Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences, 98*(2), 694–699.
Hanslmayr, S., Staresina, B. P., & Bowman, H. (2016). Oscillations and episodic memory: Addressing the synchronization/desynchronization conundrum. *Trends in Neurosciences, 39*(1), 16–25.
Hebscher, M., Barkan-Abramski, M., Goldsmith, M., Aharon-Peretz, J., & Gilboa, A. (2015). Memory, Decision-Making, and the Ventromedial Prefrontal Cortex (vmPFC): The roles of subcallosal and posterior orbitofrontal cortices in monitoring and control processes. *Cerebral Cortex*. bhv220.
Hebscher, M., & Gilboa, A. (2016). A boost of confidence: The role of the ventromedial prefrontal cortex in memory, decision-making, and schemas. *Neuropsychologia, 90*, 46–58.
Hodges, J. R., Salmon, D. P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia, 30*(4), 301–314.
Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371–375.
Johnson, M. K., O'Connor, M., & Cantor, J. (1997). Confabulation, memory deficits, and frontal dysfunction. *Brain and Cognition, 34*(2), 189–206.
Kan, I. P., Larocque, K. F., Lafleche, G., Coslett, H. B., & Verfaellie, M. (2010). Memory monitoring failure in confabulation: Evidence from the semantic illusion paradigm. *Journal of the International Neuropsychological Society: JINS, 16*(6), 1006–1017.
Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience, 14*(5), 785–794.
van Kesteren, M. T., Ruiter, D. J., Fernandez, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences, 35*(4), 211–219.
Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience and Biobehavioral Reviews, 31*(7), 1003–1016.
Kopelman, M. D. (2010). Varieties of confabulation and delusion. *Cognitive Neuropsychiatry, 15*(1), 14–37.
Kopelman, M. D. (2015). What does a comparison of the alcoholic Korsakoff syndrome and thalamic infarction tell us about thalamic amnesia? *Neuroscience and Biobehavioral Reviews, 54*, 46–56.
Kopelman, M. D., Wilson, B. A., & Baddeley, A. D. (1990). *The autobiographical memory interview (manual)*. Bury St. Edmunds, England: Thames Valley Test Company.
Kopelman, M. D., Ng, N., & Van den Brouke, O. (1997). Confabulation extending across episodic memory, personal and general semantic memory. *Cognitive Neuropsychology, 14*, 683–712.
Lambon Ralph, M. A., Ehsan, S., Baker, G. A., & Rogers, T. T. (2012). Semantic memory is impaired in patients with unilateral anterior temporal lobe resection for temporal lobe epilepsy. *Brain : a journal of neurology, 135*(Pt 1), 242–258.
Lieberman, M. D., Jarcho, J. M., & Satpute, A. B. (2004). Evidence-based and intuition-based self-knowledge: An FMRI study. *Journal of Personality and Social Psychology, 87*(4), 421–435.
Liu, Z. X., Grady, C., & Moscovitch, M. (2016). Effects of prior-knowledge on brain activation and connectivity during associative memory encoding. *Cerebral Cortex*. bhw047.
Marquine, M. J., Grilli, M. D., Rapcsak, S. Z., Kaszniak, A. W., Ryan, L., Walther, K., et al. (2016). Impaired personal trait knowledge, but spared other-person trait knowledge, in an individual with bilateral damage to the medial prefrontal cortex. *Neuropsychologia, 89*, 245–253.
McCrae, S. (2010). Intuition, insight, and the right hemisphere: Emergence of higher sociocognitive functions. *Psychology Research and Behavior Management, 3*, 1–39.
Melo, B., Winocur, G., & Moscovitch, M. (1999). False recall and false recognition: An examination of the effects of selective and combined lesions to the medial temporal lobe/diencephalon and frontal lobe structures. *Cognitive Neuropsychology, 16*, 3–5.
Mendez, M. F., Fras, I. A., Kremen, S. A., & Tsai, P. H. (2011). False reports from patients with frontotemporal dementia: Delusions or confabulations? *Behavioural Neurology, 24*(3), 237–244.
Moscovitch, M. (1989). Confabulation and the frontal systems: Strategic versus associative retrieval in neuropsychological theories of memory. In H. L. Roediger, & F. I. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of endel tulving* (pp. 133–160). Hillsdale, NJ: Lawrence Erlbaum Associates.
Moscovitch, M., & Melo, B. (1997). Strategic retrieval and the frontal lobes: Evidence from confabulation and amnesia. *Neuropsychologia, 35*(7), 1017–1034.
Moscovitch, M., & Winocur, G. (2002). The frontal cortex and working with memory. *Principles of Frontal Lobe Function*, 188–209.
Murre, J. M., Graham, K. S., & Hodges, J. R. (2001). Semantic dementia: Relevance to connectionist models of long-term memory. *Brain: A Journal of Neurology, 124*(Pt 4), 647–675.
Nahum, L., Ptak, R., Leemann, B., & Schnider, A. (2009). Disorientation, confabulation, and extinction capacity: Clues on how the brain creates reality. *Biological Psychiatry, 65*(11), 966–972.
Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current Biology: CB, 23*(17), R764–R773.
Rameson, L. T., Satpute, A. B., & Lieberman, M. D. (2010). The neural correlates of implicit and explicit self-relevant processing. *NeuroImage, 50*(2), 701–708.
Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*(9), 677–688.
Rosenbaum, R. S., Gilboa, A., & Moscovitch, M. (2014). Case studies continue to illuminate the cognitive neuroscience of memory. *Annals of the New York Academy of Sciences, 1316*, 105–133.
Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Reviews, 1*(1), 63–75.
Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Science, 15*, 467–474.
Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49*, 289–318.

Schnider, A. (2003). Spontaneous confabulation and the adaptation of thought to ongoing reality. *Nature Reviews Neuroscience, 4*(8), 662–671.
Schnider, A. (2008). *The confabulating mind: How the brain creates reality*. Oxford: Oxford University Press.
Schnider, A. (2013). Orbitofrontal reality filtering. *Frontiers in behavioral neuroscience, 7*, 67.
Schnider, A., Nahum, L., & Ptak, R. (2017). What does behavioural extinction have to do with confabulation? *Cortex, 87*, 5–15.
Schnider, A., Valenza, N., Morand, S., & Michel, C. M. (2002). Early cortical distinction between memories that pertain to ongoing reality and memories that don't. *Cerebral Cortex, 12*(1), 54–61.
Segal, Z. V. (1988). Appraisal of the self-schema construct in cognitive models of depression. *Psychological Bulletin, 103*(2), 147–162.
Serra, M., La Corte, V., Migliaccio, R., Brazzarola, M., Zannoni, I., Pradat-Diehl, P., et al. (2014). Confabulators mistake multiplicity for uniqueness. *Cortex, 58*, 239–247.
Shallice, T., & Cooper, R. P. (2012). The organisation of mind. *Cortex, 48*(10), 1366–1370.
Spalding, K. N., Jones, S. H., Duff, M. C., Tranel, D., & Warren, D. E. (2015). Investigating the neural correlates of schemas: Ventromedial prefrontal cortex is necessary for normal schematic influence on memory. *The Journal of Neuroscience, 35*(47), 15746–15751.
Thorndyke, P. W., & Yekovich, F. R. (1980). A critique of schema-based theories of human story memory. *Poetics, 9*(1), 23–49.
Turner, M. S., Cipolotti, L., Yousry, T. A., & Shallice, T. (2008). Confabulation: Damage to a specific inferior medial prefrontal system. *Cortex, 44*(6), 637–648.
Turner, M., & Coltheart, M. (2010). Confabulation and delusion: A common monitoring framework. *Cognitive Neuropsychiatry, 15*(1), 346–376.
Uretzky, S., & Gilboa, A. (2010). Knowing your lines but missing your cue: Rostral prefrontal lesions impair prospective memory cue detection, but not action-intention superiority. *Journal of Cognitive Neuroscience, 22*(12), 2745–2757.
Warren, D. E., Jones, S. H., Duff, M. C., & Tranel, D. (2014). False recall is reduced by damage to the ventromedial prefrontal cortex: Implications for understanding the neural correlates of schematic memory. *The Journal of Neuroscience, 34*(22), 7677–7682.