An Electrophysiological Dissociation between Orbitofrontal Reality Filtering and Context Source Monitoring

Aurélie Bouzerda-Wahlen[1], Louis Nahum[1], Maria Chiara Liverani[1], Adrian G. Guggisberg[1,2], and Armin Schnider[1,2]

Abstract

■ Memory influences behavior in multiple ways. One important aspect is to remember in what precise context in the past a piece of information was acquired (context source monitoring). Another important aspect is to sense whether an upcoming thought, composed of fragments of memories, refers to present reality and can be acted upon (orbitofrontal reality filtering). Whether these memory control processes share common underlying mechanisms is unknown. Failures of both have been held accountable for false memories, including confabulation. Electrophysiological and imaging studies suggest a dissociation but used very different paradigms. In this study, we juxtaposed the requirements of context source monitoring and reality filtering within a unique continuous recognition task, which healthy participants performed while high-resolution evoked potentials were recorded. The mechanisms dissociated both behaviorally and electrophysiologically: Reality filtering induced a frontal positivity, absence of a specific electrocortical configuration, and posterior medial orbitofrontal activity at 200–300 msec. Context source monitoring had no electrophysiological expression in this early period. It was slower and less accurate than reality filtering and induced a prolonged positive potential over frontal leads starting at 400 msec. The study demonstrates a hitherto unrecognized separation between orbitofrontal reality filtering and source monitoring. Whereas deficient orbitofrontal reality filtering is associated with reality confusion in thinking, the behavioral correlates of deficient source monitoring should be verified with controlled experimental exploration. ■

INTRODUCTION

Source monitoring denotes the ability to verify the precise circumstances under which a memory was acquired, that is, its temporal, spatial, and emotional context, its precise content (color, size, taste) or whether it was self-generated or not (a faculty also called reality monitoring; Mitchell & Johnson, 2009; Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981). In a classic paradigm, participants had to judge whether they had previously seen a visually presented word as a word, a picture, or not at all (Johnson, Kounios, & Nolde, 1997). This type of source identification evoked potentials that dissociated from those evoked by an old/new test at frontal electrode sites after about 400 msec. Other typical tasks tested the ability to indicate whether a visually presented word had previously been heard from a male or female voice or what type of decision one had to make while incidentally learning words (Addante, Ranganath, & Yonelinas, 2012). Evoked potentials revealed frontal old/new effects starting after 400 msec, called FN400, whose amplitude, depending on the study, varied with accuracy or confidence about the source judgment (Addante et al., 2012; Leynes & Phillips, 2008; Vallesi & Shallice, 2006; Wilding & Rugg, 1996). Source monitoring has additional electrophysiological components: a left parietal old/new effect at 600–900 msec indicating recollection (Diana, Van den Boom, Yonelinas, & Ranganath, 2010; Leynes & Phillips, 2008; Tsivilis, Otten, & Rugg, 2001) and a frontal component starting at 800 msec that has been supposed to reflect postretrieval processes (Leynes & Phillips, 2008; Tsivilis et al., 2001).

It is unclear how source monitoring relates to orbitofrontal reality filtering—the ability to sense whether an upcoming thought, emanating from the association of memories acquired in the past, relates to the present and can be acted upon, or not (Schnider, 2008, 2013). The existence of this thought control mechanism has initially been deduced from clinical observation of patients who confuse reality after brain damage: They act according to ideas that have no relation with ongoing reality (mostly, they insist on pursuing their habitual professional obligations), justify their acts with seemingly invented stories (confabulations), which are mostly composed of memory fragments from their real past, and are disoriented—when questioned, they do not know the day or place and misjudge their current obligations (Nahum, Bouzerda-Wahlen,

[1]University of Geneva, [2]University Hospital, Geneva, Switzerland

Guggisberg, Ptak, & Schnider, 2012; Schnider & Ptak, 1999; Schnider, von Däniken, & Gutbrod, 1996a). Focal lesions of all hitherto described patients involved the OFC, Area 13, or structures directly connected with it (perirhinal cortex on one side and amygdala on the other; anterior medial hypothalamus; casular genu; dorsomedial thalamic nucleus; Schnider, 2008, 2013; Nys et al., 2004; Ptak et al., 2001; Schnider, Ptak, von Däniken, & Remonda, 2000; Schnider & Ptak, 1999; Schnider, Gutbrod, Hess, & Schroth, 1996; Schnider, von Däniken, & Gutbrod, 1996b; Gentilini, De Renzi, & Crisi, 1987). We have called this disorder behaviorally spontaneous confabulation (Schnider, 2008) to distinguish it from confabulations as a pure verbal phenomenon, not associated with inappropriate behavior and disorientation (Nahum et al., 2012). These confabulations also preferentially occur after lesions in the ventromedial prefrontal area (Gilboa & Verfaellie, 2010; Schnider, 2008; Gilboa & Moscovitch, 2002).

We have found a very reliable surrogate marker for the reality confusion characterized by disorientation and behaviorally spontaneous confabulation: When such patients perform repeated runs of a continuous recognition task, always composed of the same picture set, they specifically increase their false positive rate from run to run; they increasingly believe having seen a picture within the ongoing run when indeed they have seen it in a previous run (Nahum et al., 2012; Gilboa et al., 2006; Schnider & Ptak, 1999; Schnider et al., 1996a). This difficulty is independent of recognition memory performance in the first run or the hit rate in subsequent runs (Nahum et al., 2012; Schnider & Ptak, 1999; Schnider et al., 1996a). The increase of false positives also very precisely predicts the severity of disorientation as tested with a questionnaire (Nahum et al., 2012; Schnider et al., 1996b). Recovery from reality confusion, with actions again in concordance with reality, was individually associated with recovery of the ability to control this kind of interference (Schnider, Ptak, et al., 2000). Failure in the task has additionally been described in actively hallucinating patients with schizophrenia (Badcock, Waters, Maybery, & Michie, 2005).

When healthy participants performed an adapted, more challenging version of this task while being scanned with $H_2^{15}O$ PET (which does not have the orbitofrontal artifacts typical of fMRI; Stenger, 2006), they activated the parahippocampal area in the first run of the task, when learning the pictures, but the posterior medial OFC, Area 13, in subsequent runs (Treyer, Buck, & Schnider, 2003, 2006; Schnider, Treyer, & Buck, 2000). Using evoked potential analysis in healthy participants, we found that the stimiuli on which the reality confusing patients had failed (the distracters of the ≥second run) evoked a distinct signal, unlike any other stimulus type: a frontal positivity at 200–300 msec (Schnider, Valenza, Morand, & Michel, 2002). Source analysis indicated that this frontal positivity corresponded to a significantly weaker expression of an electrocortical map configuration expressed in response to all other stimuli, characterized by extended activation of neocortical association areas (Schnider, 2003; Schnider et al., 2002). Thus, correct processing of stimuli that do not refer to present reality (distracters of Run 2 [Dis2]) induced less extended neocortical activation than all other stimuli at 200–300 msec. In light of its anatomical and behavioral specificity, we call the brain function measured with our task and obviously necessary to keep thought and behavior in phase with reality, orbitofrontal reality filtering (Schnider, 2013).

Deficient source monitoring, similar to reality filtering, has been proposed as a mechanism of confabulations (Mitchell & Johnson, 2009; Johnson et al., 1993; Johnson, 1991). Although neither the type of confabulation nor the component of source monitoring was specified, this suggestion raises the possibility that the two memory control mechanisms share commonalities. Although the evoked potential studies described above argue against this idea (source monitoring induced later potential variations than reality filtering), it has to be acknowledged that the studies used very different paradigms: Source monitoring tasks typically require the explicit recall or choice of source information about previously learned items, whereas reality filtering is tested with repeated runs of a continuous recognition tasks. For this study, we composed a continuous recognition task that combined the challenges of context source monitoring (verification of pictures' previous visual context) and reality filtering. We recorded high-density evoked potentials in healthy participants to address the following questions: (1) Does reality filtering within a task, which simultaneously challenges context source monitoring, again induce the frontal positivity and relative absence of an electrocortical map configuration at 200–300 msec that was previously observed in different reality filtering tasks (Wahlen, Nahum, Gabriel, & Schnider, 2011; Schnider et al., 2002)? (2) Does context source monitoring performed within a continuous recognition task share a common early process with reality filtering or is it expressed only after 400 msec, as in most traditional source monitoring tasks (Addante et al., 2012; Mollison & Curran, 2012; Leynes & Phillips, 2008; Vallesi & Shallice, 2006; Johnson, Kounios, et al., 1997; Wilding & Rugg, 1996)? Given the specific association of behaviorally spontaneous confabulation and disorientation with deficient reality filtering (Schnider, 2013), but apparently not with source monitoring (Johnson, O'Connor, & Cantor, 1997), we hypothesized that the two mechanisms would dissociate, even when tested within a common continuous recognition paradigm.

METHODS

Participants

Nineteen right-handed participants with no history of neurological or psychiatric illness and normal vision were paid to take part in the study. Four participants did not perform the full experiment (two of them because of

poor understanding of the task and the other two because of technical problems) and were excluded from the study, restricting the analysis to 15 participants (12 women, age = 24.9 ± 3.9 years). Participants gave written informed consent to participate in this study, which was approved by the institutional ethical committee.

Procedure and Task

The design of the task is presented in Figure 1C. Participants performed an experimental task, which was split into three independent blocks, separated by 10 min, to prevent fatigue. Each block had two runs of a continuous recognition task (Schnider, 2003), composed of the same set of 52 different pictures (Snodgrass & Vanderwart, 1980). Pictures appeared within one of two different visual contexts: (1) red and blue circles or (2) green and yellow squares.

In both runs of each block, participants had to indicate, as fast as possible, whether they had seen the same picture in the same context within the ongoing run. They pressed the left button of a response box with their right index to indicate "yes, already seen" or the right button with their right middle finger to indicate "no, not seen, yet." Stimuli were presented on a computer screen for 2000 msec; ISI was 700 msec.
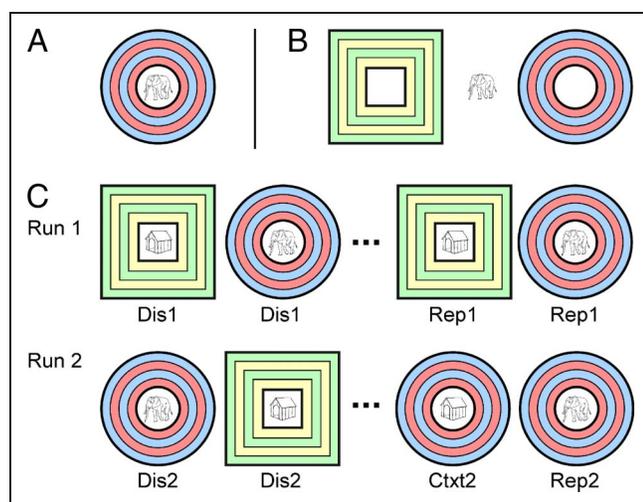

Figure 1. Task design. (A, B) Learning task preceding the main task. (A) Participants saw 12 line drawings within their correct visual context. (B) Then they were presented with each line drawing with the two contexts at their side. After their response, the drawing was again shown in its correct context, as in A. All items that had received the wrong choice of context were then repeated. (C) Main experimental task. Both runs were composed of the same items, arranged in different order. Distracters are stimuli that appear for the first time within a run (Dis1, Dis2). Repetitions are stimuli that are repeated within the same context as previously seen (Rep1, Rep2). Context change stimuli (Ctx2) are stimuli that reappeared in the ongoing run within the other visual context. The task was split into three blocks, which were composed of different pictures and were separated by a break of 10 min.

In the first runs of the three blocks, pictures always reappeared after 6–10 intervening stimuli in the same context, allowing participants to learn the association between the pictures and their context. First appearances were designated distracters of Run 1 (Dis1); recurring pictures were designated repetitions of Run 1 (Rep1).

The main variations occurred in the second runs where pictures first appeared in the same context as in the first run (Dis2; 52 stimuli) but then recurred either in the same (repetition of Run 2 [Rep2]; 26 stimuli) or the alternate context (distracters by context [Ctx2]; 26 stimuli).

Data of the three blocks were pooled for the analysis. As conditions Dis1, Dis2, and Rep1 had 52 items per block (156 items total per stimulus type) whereas conditions Ctx2 and Rep2 had 26 (78 total per stimulus type), only half of the responses to Dis1, Dis2, and Rep1 were randomly selected in each participant to enter analysis.

Main Stimuli of Interest

The processing of Dis2 stimuli was our marker for reality filtering. These are the stimuli on which reality confusing patients, who produced confabulations and were disoriented, failed; they had a specific increase of false positives in response to Dis2 stimuli (Nahum et al., 2012; Schnider, Ptak, et al., 2000; Schnider & Ptak, 1999; Schnider et al., 1996a, 1996b). Processing of these stimuli activated the posterior medial orbitofrontal cortex (Schnider, Treyer, et al., 2000) and induced a specific frontal positivity at 200–300 msec in previous evoked potential studies (Wahlen et al., 2011; Schnider et al., 2002).

Context source monitoring was studied by two parameters: (1) the difference between the first and second run, because only the second run required context monitoring, and (2) the processing of Ctx2 stimuli, which required the ability to distinguish between stimuli reoccurring in the same or a different context, in comparison with Dis2, which had not been preceded by a similar stimulus in the ongoing second run. Responses to Rep2 stimuli, which presumably also challenge context source monitoring, will also be reported although they required a "yes" answer, which might influence the evoked potential response.

Preceding Learning Task

Pilot studies had shown that participants performed at chance for context changes in the second run when they had seen the stimuli only during the preceding first continuous recognition run, that is twice as new and repeated stimuli. Therefore, before performing the experimental task, participants first learned the stimuli within their context. They learned the 156 pictures in 13 series of 12 pictures in a selective reminding task (Figure 1A, B). The 12 pictures were presented within their context, one after the other, on a computer screen for 3000 msec (ISI of 1000 msec). After the presentation of 12 stimuli, each

stimulus was presented together with the two visual contexts presented to their right and left. Participants had to press the button on the side of the correct context or the middle button if they were not sure. An auditory feedback was then delivered (pleasant sound for correct answers, unpleasant sound for incorrect answers; no sound for the "I do not know" responses), followed by presentation of the picture within its correct context. All stimuli were repeated until participants correctly recalled the picture-context association of the whole 12 pictures series. Under this condition, learning was rapid: Participants needed a second presentation for only 4.5% of the stimuli and a third presentation for 0.4%. Despite slow RTs (2703 ± 510 msec), they were self-confident about their answers; "I do not know" responses made up only 2.1% of all responses.

After the initial learning task and before the experimental task, participants had a 30-min break during which the electrodes for the EEG were installed. This interval was chosen to prevent reality filtering from being challenged in the first run of the main task. In a clinical study, healthy participants had no increase of false positives when there was only a 5-min break between the runs (Schnider & Ptak, 1999). Healthy participants felt challenged and displayed orbitofrontal activation in $H_2{}^{15}O$ PET only when the runs followed each other in rapid succession, separated by a 1-min break (Treyer et al., 2003, 2006; Schnider, Treyer, et al., 2000), but not when the runs were separated by 10 min (Schnider, 2008).

Analysis of Behavioral Data

Repeated-measures ANOVAs on RTs and correct responses were made with Stimulus type (Dis1, Rep1, Dis2, Ctx2, and Rep2) as the repeated within-subject factor. Post hoc analysis was made using Fisher's test. Planned comparisons concerned differences of RTs and response accuracy between both runs and between context change stimuli (Ctx2) and the four other conditions (Dis1, Rep1, Dis2, and Rep2).

ERP Acquisition and Preprocessing

The EEG was continuously recorded during the experimental task with an Active-Two Biosemi EEG system (Biosemi V.O.F.; Amsterdam) using 128 electrodes at a sampling rate of 512 Hz and a bandwidth filter of 0–104 Hz. Electrode impedance was kept below 20 kΩ.

Cartool software (https://sites.google.com/site/fbmlab/cartool/) was used to conduct all analyses. Epochs from −100 to 800 msec poststimulus onset were band-pass filtered to 1–30 Hz and then averaged for each condition and each participant to calculate the ERPs. ERPs were recalculated against the average reference. Baseline correction was applied by subtracting the 100-msec prestimulus period. Only correct trials were retained. Epochs with artifacts higher than ±100 μV, eye blinks, or other artifacts were rejected. Isolated artifact electrodes were interpolated using a 3-D spline interpolation (Perrin, Pernier, Bertrand, Giard, & Echallier, 1987). The final number of epochs entering the analyses is listed in Table 1.

We limited analysis to the period from 0 to 800 msec because RTs with our paradigm are so short (see Results) that effects beyond 800 msec risk to be contaminated by motor preparation and motor execution rather than memory monitoring effects. Although having the disadvantage of late motor interference beyond 800 msec, our paradigm assured that the participants really committed to the requested mental processes while attempting to respond as quickly as possible. As it was not the goal of this study to reproduce results concerning the late stages of source monitoring but to search for an early signature of content source monitoring at 200–300 msec or beyond 400 msec, this period of analysis was sufficient to answer the questions raised in the study.

Waveform Analysis

Previous studies have shown that reality filtering and recognition effects can be seized over frontal, central, and posterior electrode sites (Wahlen et al., 2011; Schnider et al., 2002), but that the precise optimal electrode position may vary. We, therefore, grouped electrodes into ROIs for the analysis: frontal, central, and posterior ROIs (as in the previous studies), to which we added a central right and central left ROIs. Each ROI consisted of 15 electrodes except for the central cluster, which was composed of 14 electrodes (Figure 2A). Within each ROI, ERPs of the corresponding electrodes were averaged for each condition and each participant. Mean trace amplitudes were subjected to repeated-measures ANOVAs with the two factors ROI (frontal, central, central right, central left, or posterior) and Stimulus type (Dis1, Rep1, Dis2, Ctx2, or Rep2) and to post hoc Fischer's tests (with a significance level of $p < .05$). Separate tests were performed for three time periods: According to the questions raised in the

Table 1. Behavioral Results

Stimulus Type	Correct Responses (%)	RT (msec)	Epochs for ERP
Dis1	90.8 ± 0.6	866.7 ± 23	54 ± 15
Rep1	88.5 ± 1.1	850.1 ± 23.1	55 ± 14
Dis2	86.2 ± 1.9	1004.7 ± 33.5	53 ± 16
Ctx2	81.7 ± 4.1	1212.6 ± 39.6	51 ± 15
Rep2	88.1 ± 2.8	1101.1 ± 41.2	53 ± 15

Values indicate mean ± standard deviation (*SD*) for correct responses, RTs, and the number of epochs per participant and stimulus type that entered the evoked potential analyses. Abbreviations: Dis1 = distracters of Run 1; Rep1 = repetitions in Run 1; Dis2 = distracters of Run 2 (measure of reality filtering); Rep2 = repetitions in Run 2 (reappearances within the same context); Ctx2 = distracters by context in Run 2 (same item reappearing in alternate context; measure of context monitoring).

Figure 2. Waveform analysis. In the bottom right of the figure, we show the arrangement of the five ROIs on a scalp (frontal, central, central left, central right, and posterior). The black dots indicate electrodes included in each ROI. Grand-averaged ERPs of each ROI in response to the five conditions (Dis1, Rep1, Dis2, Ctx2, Rep2). Repeated-measures ANOVAs applied on the mean amplitudes were performed across the two time windows boxed in gray: 200–300 msec and 400–800 msec and the third period in-between (300–400 msec). To increase the readability of the curves, the original curves were transformed to vector graphics.
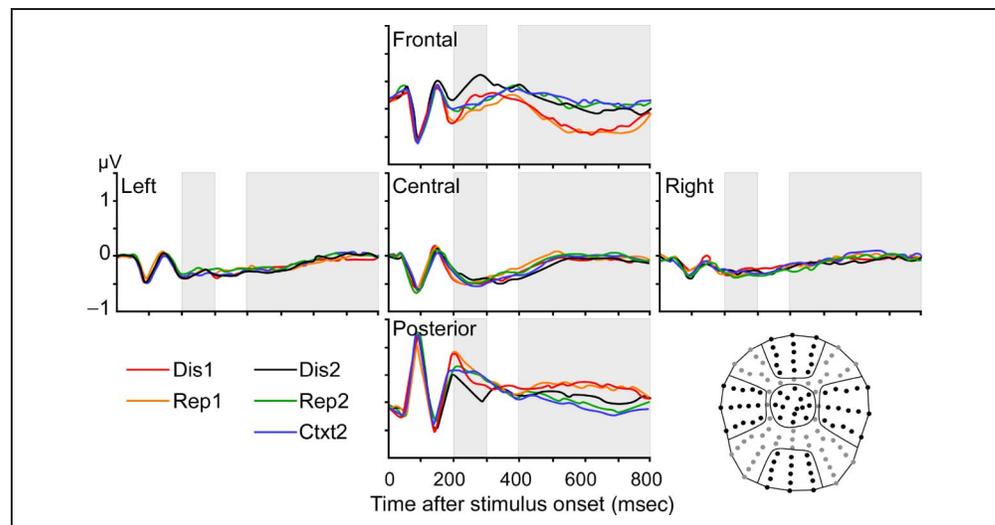
introduction, the periods of main interest were 200–300 msec and 400–800 msec. For completeness and given the novelty of the present paradigm, we separately analyzed the intermittent period of 300–400 msec.

Spatiotemporal Analysis

A modified hierarchical cluster analysis (Topographic Atomize and Agglomerate Hierarchical Clustering) was performed across the five experimental conditions (Dis1, Rep1, Dis2, Ctx2, Rep2) to determine periods of stable configuration of electrocortical voltage distribution, also called "maps," over the whole set of 128 electrodes (Brunet, Murray, & Michel, 2011; Murray, Brunet, & Michel, 2008; Michel et al., 2004). Statistical smoothing was used to eliminate temporally isolated maps with low strength (Pascual-Marqui, Michel, & Lehmann, 1995). As additional constraints, the duration of a given scalp topography had to be ≥20 msec. Clusters that correlated ≥90% were merged (Wahlen et al., 2011). The number of maps explaining the averaged data set was determined by cross validation and the Krzanowski–Lai criterion (Pascual-Marqui et al., 1995). The appearance of these maps in the individual data was then determined with a fitting procedure allowing to establish how well the maps explained individual patterns of activity (global explained variance [GEV]) and their duration (Murray et al., 2008; Michel et al., 2004). These individual measures (GEV and duration) were then subjected to repeated-measures ANOVAs with the two factors Stimulus Type and Map; post hoc comparisons used Fischer's tests. This analysis was made over the whole 800-msec time span with no predetermined temporal windows.

Source Estimation

In periods with significantly different ERPs, source estimation was made to determine the anatomical underpinnings of these differences (Murray et al., 2008; Michel et al., 2004). We used a distributed linear inverse solution based on a local autoregressive average (LAURA) model comprising a solution space of 3005 nodes (Grave de Peralta Menendez, Murray, Michel, Martuzzi, & Gonzalez Andino, 2004; Grave de Peralta Menendez, Gonzalez Andino, Lantz, Michel, & Landis, 2001). Current distribution was calculated within the gray matter of the average brain provided by the Montreal Neurological Institute. Paired t tests between the conditions of interest were then calculated for each node. Nodes with a $p < .01$ for at least 20 msec were retained as significant.

RESULTS

Behavioral Results

Accuracy and RTs in the experimental task are shown in Table 1. Repeated-measures ANOVAs revealed that Stimulus type had a significant effect on accuracy, $F(4, 56) = 3.9$, $p < .01$, and on RTs, $F(4, 56) = 70.3$, $p < .001$. Post hoc tests showed that accuracy was significantly lower in response to Ctx2 than all other types except Dis2.

RTs differed among all stimulus types except between Dis1 and Rep1, which were similar. Ctx2 stimuli evoked the slowest responses.

Planned comparisons confirmed that participants were slower, $F(1, 14) = 150.4$, $p < .001$, and less accurate, $F(1, 14) = 5.6$, $p < .05$, in response to Ctx2 items than all other stimulus types. Thus, within the second run, context source monitoring (Ctx2) required more time and effort than reality filtering (Dis2). RTs were significantly slower in the second run compared with the first run, $F(1, 14) = 88.3$, $p < .001$, but accuracy did not differ between the two runs, $F(1, 14) = 3.6$, $p = .08$.

There was no evidence of fatigue. On the contrary, RTs decreased from block to block (effect of Block,

$F(2, 28) = 25.7$, $p < .001$; post hoc testing, Block 1 > Block 2 > Block 3) and accuracy increased (effect of block, $F(2, 28) = 4.8$, $p < .05$; post hoc testing, Block 1 < Block 2 = Block 3), indicating that participants got accustomed to the difficulty of the task.

Waveform Analysis

Waveform analysis yielded results previously observed with separate reality filtering and source monitoring tasks in the two main periods of analysis and did not reveal any new commonality between the two processes. Specifically, the following results were obtained (Figure 2).

Between 200 and 300 msec, the amplitude of the ERP in response to Dis2 was distinct from the other four conditions on frontal (more positive) and posterior (more negative) ROIs. Repeated-measures ANOVA confirmed a significant interaction of ROI × Stimulus Type, $F(16, 224) = 10.6$, $p < .001$. Post hoc tests confirmed that Dis2 induced the most positive frontal and the most negative posterior waveform in comparison with all other condi-

tions, a result similar to earlier studies (Wahlen et al., 2011; Schnider et al., 2002). Ctx2 had no particular signature: It did not differ from any stimulus type other than Dis2 in this period. No difference was found in the three central ROIs (central, central left, and central right) within this time window.

Between 400 and 800 msec, there was a strong difference between the runs: Stimuli of the first run (Dis1, Rep1) induced the most negative frontal and the most positive posterior waveforms, whereas those of the second run (Dis2, Ctx2, and Rep2) induced the most positive frontal and the most negative posterior waveforms. The difference was confirmed by repeated-measures ANOVAs, $F(16, 224) = 11.6$, $p < .001$. Furthermore, post hoc testing also revealed a difference among the conditions of the second run in the posterior ROI: Dis2 induced a more positive waveform than Ctx2. There was no difference in the three central ROIs (central, central left, and central right) within this time widow.

In the intermediate period from 300 to 400 msec, for which we had no specific hypothesis, there was a

Figure 3. Spatiotemporal analysis. (A) Temporal distribution of the nine cortical maps obtained from segmentation of the grand mean ERPs over 800 msec. Red indicates positivity, blue indicates negativity. Maps that differed between task conditions are boxed. (B) Distribution of stable map configurations over 800 msec in the five conditions. The numbers below each segment indicate the most representative map of this period. The amplitude of the curves reflects the GFP. Segments in color indicate maps with significant differences between conditions. (C) Results of post hoc tests of the repeated-measures ANOVAs for Maps 3 and 7 regarding GEV and duration of the map. Horizontal lines indicate significant differences between two conditions.

significant interaction between the two factors (ROI and Stimulus type; $F(16, 224) = 1.8, p < .05$), which reflected an attenuated perpetuation of the effects seen at 200–300 msec up to 350 msec. In the frontal ROI, Dis2 induced a more positive deflection than Dis1, Rep1, and Ctx2 but did not differ from Rep2. No differences were present in the other ROIs.

Spatiotemporal Analysis

Figure 3 shows the spatiotemporal analysis of the grand mean ERPs of the five conditions. Nine different cluster maps were identified over 800 msec (Figure 3A). Figure 3B shows the temporal succession and the strength of the maps that dominated at the different points in time over all participants (individual participants may express different degrees of these maps or different maps for various durations).

Between 170 and 290 msec, repeated-measures ANOVAs with the two factors Stimulus type (five types) and Maps present in this time window (Maps 3 and 4; Figure 3A) yielded a significant interaction of GEV, $F(4, 56) = 7.1$, $p < .001$, and duration of Maps, $F(4, 56) = 3.1, p < .05$. Post hoc test (Figure 3C) showed that Map 3 was less present and shorter in response to Dis2 than in response to the other conditions (significant differences are shown in Figure 3C). In contrast, Map 3 was significantly more present in Rep1 than in Dis2 and Rep2.

At 420–800 msec, electrocortical maps differed more between runs than stimulus types (Figure 3B). Repeated-measures ANOVAs with Stimulus type and Maps (4, 5, 6, 7, 8, and 9) as factors showed a significant interaction of GEV, $F(20, 280) = 2.6, p < .001$, and duration of Maps, $F(20, 280) = 3.6, p < .001$. Post hoc analysis revealed that only the cortical Map 7 significantly differed among the conditions: It was more present and longer in response to conditions of the first run (Dis1 and Rep1) than the second run (Dis2, Ctx2, Rep2). Map 7 was less present in stimulus type Ctx2 than in all other types. Statistical differences are shown in Figure 3C.

This analysis thus replicates earlier findings concerning reality filtering (significantly weaker expression of a map at approximately 200–300 msec in response to Dis2; Wahlen et al., 2011; Schnider et al., 2002). In addition, it indicates that the waveform differences observed beyond 400 msec, specific to context source monitoring, reflect at least partially different electrocortical map configurations rather than only amplitude modulations of the same configuration, thus implying different processes (Murray et al., 2008; Michel et al., 2001).

Source Estimation

According to the main questions of this study and in concordance with the results of the waveform and spatiotemporal analyses, source estimation analysis was con-

Figure 4. Sources estimation. Areas with significantly different current densities as determined from group-averaged source estimations during the two time periods. Red and blue areas depict solution points with statistically significant differences ($p < .01$ for ≥20 msec). (A) 250–300 msec for the contrast Dis2–Ctx2. Red areas indicate stronger activation in response to Dis2, thus reflecting superior activation during reality filtering; blue areas indicate stronger activation in response to Ctx2, thus weaker activation during reality filtering. (B) 500–800 msec for Run 1–Run 2. Red areas indicate stronger activation in response to stimuli of Run 1, thus during the initial learning of the stimulus–context associations. There was no area of superior activation during content source monitoring, that is, in the comparison Run 2–Run 1.

ducted in the period from 200 to 300 msec to compare activation induced by Dis2 versus Ctx2 stimuli and from 400 to 800 msec to compare activation over all stimuli of the first run versus the second run. Within these windows, the following differences in activation were found:

- From 250 to 300 msec, Dis2 more strongly activated the posterior medial orbitofrontal (ventromedial prefronal) area than Ctx2 (red in Figure 4A) and less strongly (or Ctx2 more strongly) the bilateral retrosplenial area and right superior parietal cortex (blue in Figure 4A). As Ctx2 did not differ from the other stimuli, except for Dis2, in the waveform and spatiotemporal analysis in this period, it can be considered representative of these other stimulus types.
- From 500 to 800 msec, Run 1 activated two regions (red in Figure 4B) more strongly than Run 2: the left posterior parahippocampal area and the superior parietal cortex. No increased activation was found in Run 2 in comparison with Run 1.

DISCUSSION

This study demonstrates that orbitofrontal reality filtering dissociates from context source monitoring, even when both mechanisms are juxtaposed within a unique continuous recognition paradigm combining their challenges. Despite the relative complexity of the task, both mechanisms displayed the signatures known to them from tasks exploring them separately: Reality filtering was expressed at approximately 200–300 msec, content source monitoring from 400 msec on. There was no distinct common processing period for the two functions. Thus, the different results described in earlier studies cannot be

attributed to different task designs but appear to reflect a true functional dissociation between orbitofrontal reality filtering and content source monitoring.

On the behavioral level, both the acquisition and the subsequent processing of source information varied between context source monitoring and reality filtering. The acquisition of context information required a conscious effort. Pilot experiments had shown that, when participants saw the stimuli only once during a continuous recognition run, their subsequent recognition of the context was random. An explicit learning task was necessary for them to acquire contextual information. This is different from the information being processed in reality filtering. The fact that participants have encountered information only once in a previous run of a continuous recognition task is sufficient to activate reality filtering, although normal reality filtering appears to be a very rapid and efficacious process in the healthy brain whose challenge rapidly decreases with time: Healthy participants need to make runs in rapid succession, as in the present experiment, to feel somewhat challenged and to activate the posterior medial OFC (Treyer et al., 2003; Schnider, Treyer, et al., 2000). When runs are separated by 5 min, healthy participants have no measurable difficulty with this task (Schnider & Ptak, 1999). In contrast, patients with reality confusion after orbitofrontal damage may have an increase from the second run on, even when it is made an hour or more after the first run (Nahum et al., 2012; Schnider & Ptak, 1999; Schnider et al., 1996a, 1996b). Also, in healthy participants, reality filtering as measured with repeated continuous recognition tasks appears to be an intuitive process: In previous studies, performance remained virtually perfect and RTs did not increase from the first run to the second run (Wahlen et al., 2011; Schnider et al., 2002). This was different in this study, in which the second run demanded context source monitoring in addition to reality filtering: RTs were markedly slower in the second run. Even in this setting, stimuli demanding primarily reality filtering (Dis2) were significantly faster and more accurately processed than stimuli demanding context source monitoring (Ctx2). Thus, both the acquisition and the monitoring of contextual source information dissociate from orbitofrontal reality filtering; recollecting the context in which a memory was acquired is more effortful and takes more time than recognizing whether a memory pertains to present reality or not.

On the electrophysiological level, the two processes dissociated. At 200–300 msec, stimuli requiring reality filtering (Dis2) did not induce the frontal negativity induced by all other stimuli and produced a weaker expression of a distinct electrocortical map configuration than all other stimuli. This result agrees with earlier studies examining reality filtering in comparison with recognition memory (Schnider et al., 2002) or in conjunction with strategic content monitoring (Wahlen et al., 2011). This study shows that this early signal comes up when-

ever reality filtering is required, irrespective of concomitant task demands. This observation underscores the idea that this signal reflects the process of orbitofrontal reality filtering.

Source analysis indicated a distinct anatomical correlate of this process (Figure 4A). At 250–300 msec, the processing of Dis2 stimuli induced stronger activation of the posterior medial orbitofrontal area and less activation of neocortical association areas (retrosplenial and superior parietal cortex) in comparison with Ctx2 stimuli. This finding gives statistical support to a hypothesis previously formulated on the basis of simple source analyses, with no statistical validation (Schnider, 2003). The idea is that any upcoming thought (memory) that does not relate to present reality—a fantasy—induces an orbitofrontal signal at 200–300 msec, which prevents widespread neocortical activation from occurring. By the time the content of a thought (memory) is consciously recognized, which appears to happen around 400–600 msec (Schnider et al., 2002), the OFC has already signaled whether it relates to ongoing reality, or not, and adapted its cortical format accordingly (Schnider, 2008, 2013).

We hypothesized (Schnider, 2008, 2013) that this signal corresponds to the one emitted by select orbitofrontal neurons (particularly dense in Area 13) when an anticipated reward (outcome) does not occur (Rosenkilde, Bauer, & Fuster, 1981). Such events are further signaled by transient decreases of the firing rate of dopaminergic neurons in the midbrain (Schultz, Dayan, & Montague, 1997). Many observations support this hypothesis: In amnesic participants, the inability to process the nonoccurrence of anticipated outcomes (extinction trials) is tightly associated with reality confusion as evident in disorientation and behaviorally spontaneous confabulation (Nahum, Ptak, Leemann, & Schnider, 2009). In healthy participants, such trials evoke a frontal positivity at 200–300 msec (Schnider, Mohr, Morand, & Michel, 2007), which emanates from the posterior medial OFC (Nahum, Gabriel, & Schnider, 2011; Schnider, Treyer, & Buck, 2005). In agreement with the dopamine hypothesis of reward processing, we found that reality filtering as measured with different versions of the present task activated subcortical structures known to participate in the reward system (Treyer et al., 2003) and was modified by medication modulating dopaminergic transmission (Schnider, Guggisberg, Nahum, Gabriel, & Morand, 2010; Pihan, Gutbrod, Baas, & Schnider, 2004). Thus, orbitofrontal reality filtering constitutes a distinct memory control mechanism with a defined behavioral correlate, whose physiological and pharmacological basis is increasingly better understood.

Context source monitoring had no specific electrophysiological expression at 200–300 msec, neither in this study nor in previous ones using different paradigms (Addante et al., 2012; Mollison & Curran, 2012; Leynes & Phillips, 2008; Vallesi & Shallice, 2006; Johnson, Kounios, et al., 1997; Wilding & Rugg, 1996). However, it had a

distinct signature from 400 msec on: a prolonged frontal positivity (posterior negativity) in response to all stimuli of the second run as compared with those of the first run. This potential has not been observed in pure reality filtering tasks (Wahlen et al., 2011; Schnider et al., 2002) but has repeatedly been described in source monitoring tasks using various paradigms, where it was interpreted as an old/new effect between stimuli on which a source judgment had previously been made and new items (Addante et al., 2012; Hayama, Johnson, & Rugg, 2008; Leynes & Phillips, 2008; Vallesi & Shallice, 2006; Johnson, Kounios, et al., 1997; Wilding & Rugg, 1996).

It is likely that this relatively late potential observed in this study reflects elaborative source processing as it was not described in incidental source changes. Tsivilis et al. (2001) tested recognition of previously seen objects presented on distinct backgrounds. Participants indicated familiarity with the object, irrespective of the background. The sole recomposition of known objects with known backgrounds—comparable to the procedure used in this study—evoked no distinct potential change as compared with the presentation of previously seen object–background combinations. Thus, incidental, task-irrelevant changes of context did not induce potential changes as observed in our task and explicit source monitoring tasks.

The frontal positivity after 400 msec seems to be specific for context source monitoring rather than the monitoring of overall stimulus alterations induced by the alternate context: In an earlier study comparing reality filtering with strategic content monitoring, as proposed by Gilboa et al. (2006) and Moscovitch and Melo (1997), the presentation of stimuli that only resembled but were not identical with previously seen pictures induced a frontal negativity at 200–300 msec, which emanated from particularly extended activation of visual association cortex (Wahlen et al., 2011), presumably reflecting a visual comparision process. Thus, content monitoring was expressed in the same period as reality filtering, but with inverse polarity based on activation of different brain areas. In summary, context source monitoring is distinct not only from orbitofrontal reality filtering but also from strategic monitoring of memories' content.

Source analysis did not reveal any circumscribed brain activation characteristic of context source monitoring. Whereas the first run induced significantly stronger parahippocampal and parietal activation at 500–800 msec (Figure 4B), probably reflecting learning of the stimulus–context associations, Run 2 (context monitoring) induced no increased activation in any region. According to the RTs observed in this study, context source monitoring is a relatively slow elaborative process, even when participants are pushed to respond as quickly as possible. Indeed, studies that let participants mentally elaborate on source information for up to 2 sec before demanding a response suggested that source information is processed well beyond the 800 msec analyzed in this study (Addante et al., 2012; Diana et al., 2010; Johnson, Kounios, et al., 1997). One may speculate that an elaborative process like context source monitoring invokes varying brain areas in irregular succession, depending on the advancement of the monitoring process, so that there is no precise synchronicity of brain areas over multiple stimuli; source estimation on the basis of electroencephalography might have a too high temporal resolution to seize such an inconsistent succession of mental processes. Functional magnetic resonance imaging, with its low temporal resolution, indicated participation of multiple brain areas in source monitoring, including the medial-temporal lobe (Ross & Slotnick, 2008; Peters, Suchan, Koster, & Daum, 2007), precuneus (Lundstrom, Ingvar, & Petersson, 2005), and prefrontal cortex (Mitchell & Johnson, 2009; Mitchell, Johnson, Raye, & Greene, 2004). This pattern is distinct from the orbitofrontal activation associated with reality filtering (Treyer et al., 2006; Schnider, Treyer, et al., 2000).

This study, in conjunction with previous studies separating orbitofrontal reality filtering from content monitoring (Bouzerda-Wahlen, Nahum, Ptak, & Schnider, 2013; Wahlen et al., 2011), underscores the existence of diverse memory monitoring mechanisms which presumably dissociate from each other on the behavioral, anatomical, and physiological level. The two mechanisms juxtaposed in this study have been tacitly linked by the claim that both may explain confabulations. Deficient orbitofrontal reality filtering has been shown to induce reality confusion, evident in disorientation and confabulation (Nahum et al., 2012; Schnider & Ptak, 1999; Schnider et al., 1996a). The link between confabulation and source monitoring has not received empirical support yet (Johnson, O'Connor, et al., 1997). Although this study does not question the importance of source monitoring in everyday life, it makes it highly unlikely that this function might explain similar mental failures or faculties as orbitofrontal reality filtering. The behavioral correlates of deficient source monitoring should be verified with controlled experimental exploration.

Acknowledgments

REFERENCES

Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: Evidence of accurate source recognition without recollection. *Neuroimage, 62,* 439–450.

Badcock, J. C., Waters, F. A., Maybery, M. T., & Michie, P. T. (2005). Auditory hallucinations: Failure to inhibit irrelevant memories. *Cognitive Neuropsychiatry, 10,* 125–136.

Bouzerda-Wahlen, A., Nahum, L., Ptak, R., & Schnider, A. (2013). Mechanism of disorientation: Reality filtering versus content monitoring. *Cortex, 49,* 2628–2636.

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. *Computational Intelligence and Neuroscience, 2011,* 813–870.

Diana, R. A., Van den Boom, W., Yonelinas, A. P., & Ranganath, C. (2010). ERP correlates of source memory: Unitized source information increases familiarity-based retrieval. *Brain Research, 1367,* 278–286.

Gentilini, M., De Renzi, E., & Crisi, G. (1987). Bilateral paramedian thalamic artery infarcts: Report of eight cases. *Journal of Neurology, Neurosurgery and Psychiatry, 50,* 900–909.

Gilboa, A., Alain, C., Stuss, D. T., Melo, B., Miller, S., & Moscovitch, M. (2006). Mechanisms of spontaneous confabulations: A strategic retrieval account. *Brain, 129,* 1399–1414.

Gilboa, A., & Moscovitch, M. (2002). The cognitive neuroscience of confabulation: A review and a model. In A. Baddeley, M. D. Kopelman, & B. Wilson (Eds.), *The handbook of memory disorders* (2nd ed., pp. 315–342). West Sussex, UK: Wiley.

Gilboa, A., & Verfaellie, M. (2010). Telling it like it isn't: The cognitive neuroscience of confabulation. *Journal of the International Neuropsychological Society, 16,* 961–966.

Grave de Peralta Menendez, R., Gonzalez Andino, S., Lantz, G., Michel, C. M., & Landis, T. (2001). Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. *Brain Topography, 14,* 131–137.

Grave de Peralta Menendez, R., Murray, M. M., Michel, C. M., Martuzzi, R., & Gonzalez Andino, S. L. (2004). Electrical neuroimaging based on biophysical constraints. *Neuroimage, 21,* 527–539.

Hayama, H. R., Johnson, J. D., & Rugg, M. D. (2008). The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: Specific or non-specific? *Neuropsychologia, 46,* 1211–1223.

Johnson, M. K. (1991). Reality monitoring: Evidence form confabulation in organic brain disease patients. In G. P. Prigatano & D. L. Schacter (Eds.), *Awareness of deficit after brain injury. Clinical and theoretical issues* (pp. 176–197). New York: Oxford University Press.

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

Johnson, M. K., Kounios, J., & Nolde, S. F. (1997). Electrophysiological brain activity and memory source monitoring. *NeuroReport, 8,* 1317–1320.

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

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

Leynes, P. A., & Phillips, M. C. (2008). Event-related potential (ERP) evidence for varied recollection during source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 741–751.

Lundstrom, B. N., Ingvar, M., & Petersson, K. M. (2005). The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. *Neuroimage, 27,* 824–834.

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology, 115,* 2195–2222.

Michel, C. M., Thut, G., Morand, S., Khateb, A., Pegna, A. J., Grave de Peralta, R., et al. (2001). Electric source imaging of human brain functions. *Brain Research: Brain Research Reviews, 36,* 108–118.

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

Mitchell, K. J., Johnson, M. K., Raye, C. L., & Greene, E. J. (2004). Prefrontal cortex activity associated with source monitoring in a working memory task. *Journal of Cognitive Neuroscience, 16,* 921–934.

Mollison, M. V., & Curran, T. (2012). Familiarity in source memory. *Neuropsychologia, 50,* 2546–2565.

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

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography, 20,* 249–264.

Nahum, L., Bouzerda-Wahlen, A., Guggisberg, A., Ptak, R., & Schnider, A. (2012). Forms of confabulation: Dissociations and associations. *Neuropsychologia, 50,* 2524–2534.

Nahum, L., Gabriel, D., & Schnider, A. (2011). Human processing of behaviorally relevant and irrelevant absence of expected rewards: A high-resolution ERP study. *PLoS One, 6,* e16173.

Nahum, L., Ptak, R., Leemann, B., & Schnider, A. (2009). Disorientation, confabulation, and extinction capacity. Clues on how the brain creates reality. *Biological Psychiatry, 65,* 966–972.

Nys, G. M., van Zandvoort, M. J., Roks, G., Kappelle, L. J., de Kort, P. L., & de Haan, E. H. (2004). The role of executive functioning in spontaneous confabulation. *Cognitive and Behavioral Neurology, 17,* 213–218.

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: Model estimation and validation. *IEEE Transactions on Biomedical Engineering, 42,* 658–665.

Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology, 66,* 75–81.

Peters, J., Suchan, B., Koster, O., & Daum, I. (2007). Domain-specific retrieval of source information in the medial temporal lobe. *European Journal of Neuroscience, 26,* 1333–1343.

Pihan, H., Gutbrod, K., Baas, U., & Schnider, A. (2004). Dopamine inhibition and the adaptation of behavior to ongoing reality. *NeuroReport, 15,* 709–712.

Ptak, R., Birtoli, B., Imboden, H., Hauser, C., Weis, J., & Schnider, A. (2001). Hypothalamic amnesia with spontaneous confabulations: A clinicopathologic study. *Neurology, 56,* 1597–1600.

Rosenkilde, C. E., Bauer, R. H., & Fuster, J. M. (1981). Single cell activity in ventral prefrontal cortex of behaving monkeys. *Brain Research, 209,* 375–394.

Ross, R. S., & Slotnick, S. D. (2008). The hippocampus is preferentially associated with memory for spatial context. *Journal of Cognitive Neuroscience, 20,* 432–446.

Schnider, A. (2003). Spontaneous confabulation and the adaptation of thought to ongoing reality. *Nature Reviews Neuroscience, 4,* 662–671.

Schnider, A. (2008). *The confabulating mind. How the brain creates reality*. Oxford, UK: Oxford University Press.

Schnider, A. (2013). Orbitofrontal reality filtering. *Frontiers in Behavioral Neuroscience, 7,* 67.

Schnider, A., Guggisberg, A., Nahum, L, Gabriel, D., & Morand, S. (2010). Dopaminergic modulation of rapid reality adaptation in thinking. *Neuroscience, 167,* 583–587.

Schnider, A., Gutbrod, K., Hess, C. W., & Schroth, G. (1996). Memory without context. Amnesia with confabulations following right capsular genu infarction. *Journal of Neurology, Neurosurgery and Psychiatry, 61,* 186–193.

Schnider, A., Mohr, C., Morand, S., & Michel, C. M. (2007). Early cortical response to behaviorally relevant absence of anticipated outcomes: A human event-related potential study. *Neuroimage, 35,* 1348–1355.

Schnider, A., & Ptak, R. (1999). Spontaneous confabulators fail to suppress currently irrelevant memory traces. *Nature Neuroscience, 2,* 677–681.

Schnider, A., Ptak, R., von Däniken, C., & Remonda, L. (2000). Recovery from spontaneous confabulations parallels recovery of temporal confusion in memory. *Neurology, 55,* 74–83.

Schnider, A., Treyer, V., & Buck, A. (2000). Selection of currently relevant memories by the human posterior medial orbitofrontal cortex. *Journal of Neuroscience, 20,* 5880–5884.

Schnider, A., Treyer, V., & Buck, A. (2005). The human orbitofrontal cortex monitors outcomes even when no reward is at stake. *Neuropsychologia, 43,* 316–323.

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,* 54–61.

Schnider, A., von Däniken, C., & Gutbrod, K. (1996a). The mechanisms of spontaneous and provoked confabulations. *Brain, 119,* 1365–1375.

Schnider, A., von Däniken, C., & Gutbrod, K. (1996b). Disorientation in amnesia. A confusion of memory traces. *Brain, 119,* 1627–1632.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Stenger, V. A. (2006). Technical considerations for BOLD fMRI of the orbitofrontal cortex. In D. H. Zald & S. L. Rauch (Eds.), *The orbitofrontal cortex* (pp. 423–446). Oxford, UK: Oxford University Press.

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

Treyer, V., Buck, A., & Schnider, A. (2006). Selection of currently relevant words: An auditory verbal memory study using positron emission tomography. *NeuroReport, 17,* 323–327.

Tsivilis, D., Otten, L. J., & Rugg, M. D. (2001). Context effects on the neural correlates of recognition memory: An electrophysiological study. *Neuron, 31,* 497–505.

Vallesi, A., & Shallice, T. (2006). Prefrontal involvement in source memory: An electrophysiological investigation of accounts concerning confidence and accuracy. *Brain Research, 1124,* 111–125.

Wahlen, A., Nahum, L., Gabriel, D., & Schnider, A. (2011). Fake or fantasy: Rapid dissociation between strategic content monitoring and reality filtering in human memory. *Cerebral Cortex, 21,* 2589–2598.

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