Early Cortical Distinction between Memories that Pertain to Ongoing Reality and Memories that Don't

Armin Schnider[1], Nathalie Valenza[1], Stephanie Morand[1,2] and Christoph M. Michel[2]

[1]Rehabilitation Clinic and [2]Functional Brain Mapping Laboratory, Neurology Clinic, University Hospital, Geneva, Switzerland

Patients with anterior limbic lesions, in particular of the posterior orbitofrontal cortex, often act on the basis of memories that do not relate to ongoing reality and justify their behavior with invented stories that can mostly be traced back to real events (spontaneous confabulation). Recent studies demonstrated that the patients fail to suppress activated memory traces that do not pertain to ongoing reality. In the present study, we used a similar paradigm and high-resolution event-related potentials to explore when this suppression happens. Healthy subjects made two runs of a continuous recognition task, composed of the same set of pictures, and were requested to indicate picture recurrences only within the ongoing run. Thus, performance in the first run depends on new learning, whereas the second run requires the ability to realize whether a picture is solely familiar from its occurrence in the previous run ('distracter') or whether it has already appeared in the ongoing second run ('target'). We find that correct rejection (suppression) of currently irrelevant pictures (distracters of run 2) is associated with absent negative deflection of a frontal potential and absence of a specific cortical potential map configuration after 220–300 ms. By contrast, learning and recognition of repeatedly presented information is associated with cortical amplitude modulation after 400–480 ms. These findings indicate that by the time the content of a mental association is recognized and consolidated, its cortical representation has already been adjusted according to whether it relates to ongoing reality or not. This sequence may also explain the ability to distinguish between the memory of a true event and the memory of a thought.

Introduction

Memories are the essence of thoughts, give meaning to percepts, and are the basis for ideas and plans. There is strong evidence that memories are represented in the connective pattern of neuronal networks in the cortex, ready to be activated by external stimuli or internal states (Damasio, 1989; Fuster, 1995; Singer, 1998). It is a great mystery how the brain distinguishes, among all mental associations possible at any moment, between memories that pertain to ongoing reality and memories having no relation with reality (fantasies).

Patients with anterior limbic damage, in particular damage or disconnection of the orbitofrontal cortex (OFC), may fail to adapt their thinking and behavior to ongoing reality. They consider themselves in a temporally displaced, past reality – they are disoriented – and act according to memories which may have justly guided their behavior in the past, but which do not refer to ongoing reality. For example, a woman hospitalized following rupture of an anterior communicating artery aneurysm was convinced that she was at home and had to feed her baby; but her 'baby' was over 30 years old at the time (Schnider *et al.*, 1996b). A tax accountant with extensive traumatic destruction of the OFC inadvertently left the hospital in the conviction that he had a meeting with the county's financial director (Schnider *et al.*, 1996b, 2000a). The patients justify their acts with stories

that disregard the present reality of their hospitalization and which may appear to be entirely invented (spontaneous confabulations); nonetheless, the stories can virtually always be traced back to elements of real events (Korsakow, 1892; DeLuca and Cicerone, 1991; Dalla Barba, 1993; Schnider *et al.*, 1996a,b; Kopelman *et al.*, 1997a; Burgess and McNeil, 1999; Ptak and Schnider, 1999; Schnider and Ptak, 1999). The patients' absolute conviction that what they say and do is appropriate indicates that their true problem is a confusion of currently irrelevant memories with ongoing reality; the confabulations are just the (honest) verbal manifestation of this confusion.

Our studies revealed a specific memory failure in spontaneous confabulators: when the patients made two runs of a continuous recognition task, in which they were asked to indicate recurrences of pictures, they performed similarly to non-confabulating amnesics in the first run, indicating that the two groups had similar difficulty in learning the information (Schnider *et al.*, 1996b). By contrast, when they were again presented with the same picture series – arranged in different order – 1 h later and asked to forget that they had already seen all pictures and to indicate picture recurrences only within the current run, spontaneous confabulators had a specific increase of false positive responses; they thought much more often than non-confabulating amnesics or healthy controls that they had already seen a picture in the current second run, although they had only seen it in the previous first run, 1 h previously (Schnider *et al.*, 1996b; Schnider and Ptak, 1999). That is, the patients failed to suppress the currently irrelevant memory of the pictures' previous occurrence in the first run.

The pertinence of this result is underscored by additional observations: all patients failing in the suppression demand of the second run of this task had the reality confusion typical of spontaneous confabulators; no single non-confabulating amnesic failed in this capacity (Schnider *et al.*, 1996b; Ptak and Schnider, 1999; Schnider and Ptak, 1999); particularly bizarre confabulations, indicating that a patient failed to suppress even distant mental associations, were associated with particularly severe failure in the task (Schnider and Ptak, 1999); the course of this failure precisely paralleled the course of spontaneous confabulation in all patients: recovery was always accompanied by recovery of this suppression capacity, continued confabulation by continued failure in the task (Schnider *et al.*, 2000a). This suppression failure, rather than the capacity to learn new information (as measured by the first run of the task), highly correlated with orientation regarding time, place and circumstances, a finding indicating that the spontaneous confabulators' disorientation results from the same failure (Schnider *et al.*, 1996c). It thus appears that the task does indeed measure a capacity essential for the ability to refer thinking to current reality, i.e. for monitoring of ongoing reality. Specifically, it appears that the ability to select from memory the information

that pertains to ongoing reality is reflected by the ability to recognize in the second run that an item which appears familiar only from presentation in the first run has not yet been presented in the ongoing second run. That is, processing (suppression) of the distracters of the second run appears to best mirror the anterior limbic process monitoring ongoing reality, which is defective in spontaneous confabulation.

Failure of this suppression capacity has strong anatomical specificity: whereas isolated failure in the first (learning) run was associated with medial temporal (hippocampal and parahippocampal) and neocortical (often dorsolateral prefrontal) damage, failure of suppression was exclusively seen in patients with anterior limbic damage. Most lesions involved the medial OFC itself or the basal forebrain (Schnider *et al.*, 1996b,c, 2000a; Schnider and Ptak, 1999). Single cases had lesions involving the projections to the OFC in the right capsular genu (Schnider *et al.*, 1996a), the amygdala on one side and the perirhinal cortex on the other side (Schnider *et al.*, 1996b, 2000a), or the medial hypothalamus (Ptak *et al.*, 2001). All of these structures have direct connections with the posterior medial OFC.

The absolute conviction of patients with anterior limbic damage that their imagined reality is true suggests that the suppression of currently irrelevant memories is a pre-conscious mechanism intervening before the content of a memory is recognized and consolidated. In the present study, we used high-resolution event-related potential (ERP) recordings and spatiotemporal analysis techniques to test this conjecture. The technique allows us to study the temporal dynamics of cortical network activity in the millisecond range. Based on our previous experience with an imaging study, we used a more difficult version of the paradigm than the one used in our clinical studies. Whereas spontaneous confabulators had failed to suppress distracters of the second run even when the runs were separated by up to 1 h (Schnider *et al.*, 1996b; Schnider and Ptak, 1999), a study using positron emission tomography (PET) in healthy subjects indicated that the task has to be more difficult in order to challenge the anterior limbic suppression system: when runs were made in rapid succession, separated only by 60 s, there was circumscribed activation of the posterior medial OFC (the area typically damaged or disconnected in confabulators) in the second run. As expected, the first run (learning and recognition) induced parahippocampal activation (Schnider *et al.*, 2000b). Our hypothesis for the present study was that suppression of currently irrelevant memory traces, i.e. processing of distracters in the second run, would be reflected in distinct changes of electrical cortical activity before the stage of learning and recognition.

Materials and Methods
Thirteen paid, right-handed students (eight men, five women), aged 22.6 ± 3.6 years, participated in the study. They gave written informed consent. The study was approved by the local ethical committee.

Experimental Procedure
We used an adapted version of the continuous recognition task used in our previous clinical and imaging studies (Schnider *et al.*, 1996b,c, 2000b; Schnider and Ptak, 1999). The task consists of two runs of a continuous recognition test composed from the same set of line drawings (Snodgrass and Vanderwart, 1980), arranged in different order in the two runs (Fig. 1).

In order to prevent fatigue, the task was actually split into two blocks, both having a first and a second run with 84 pictures (Fig. 1). A series of 84 pictures was composed of 48 different pictures, of which 21 were selected to reappear once (6 pictures) or twice (15 pictures) as a target during the run [36 targets (repeated presentations); 48 distracters (first presentations) per run per block]. Stimuli were presented on a computer
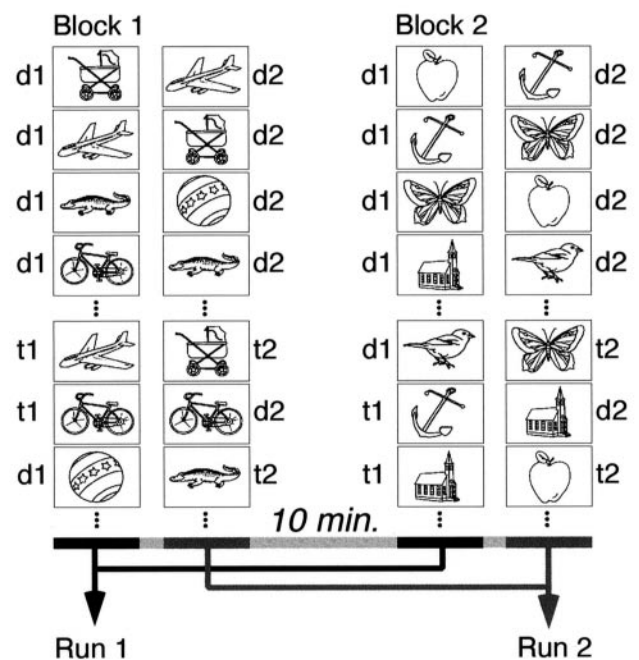

Figure 1. Design of the task. Each block is composed of a separate series of pictures. Both runs of a block are composed of the same pictures, arranged in different order. Targets are picture repetitions within a run (t1, t2 = targets of run 1 or 2), distracters are items that appear for the first time within a run (d1, d2 = distracters of run 1 or 2). Data of the two blocks were pooled to constitute run 1 and run 2.

screen for 2000 ms; interstimulus interval was 700 ms. Subjects were asked to indicate picture repetitions (targets) as fast as possible by pressing the left arrow key with their right index finger to indicate that, 'yes, the picture has already appeared within this run.' In response to first picture presentations (distracters), they were asked to press the down arrow key with their middle finger to indicate that, 'no, the picture has not appeared within this run, yet.'

The second run of each block was composed of the same picture set as the first run of the respective block, only the order of picture presentation was changed with different pictures selected to be repeated (targets) within the run (Fig. 1, 'Run 2'). The run was started 60 s after the end of the first run of the respective block. The subjects were asked to forget that they had already seen all pictures and to indicate picture recurrences only within the current run. Thus, the second run had exactly the same design and was composed of precisely the same pictures as the first run. However, it differed from the first run in that all items were already known. Subjects had a 10 min break between the two blocks. Data from the two blocks were pooled in the analysis, yielding a total of 168 responses per run (72 targets, 96 distracters).

In the first run, all pictures are initially new; a picture appearing familiar can be assumed to be a target. Thus, this run measures the ability to learn and recognize new information. Patients with amnesia following damage of the hippocampal area typically fail in this first run (Schnider *et al.*, 1996b,c; Schnider and Ptak, 1999). Healthy subjects performing the first run of this task while being scanned with PET showed activation of the parahippocampal gyrus on both sides (Schnider *et al.*, 2000b).

In the second run, all pictures are already known; responses based on familiarity alone are no longer correct. The second run demands the ability to distinguish between the currently relevant memory of an item's previous occurrence in the second run (the ongoing reality) and the currently irrelevant memory of an item's previous occurrence in the first run. All spontaneous confabulators involved in our clinical studies had increasing difficulty in this second run: they had had a specific increase of false positive responses over the first run, indicating that they failed to suppress the interference of their currently irrelevant memory of the first run (Schnider *et al.*, 1996b; Schnider and Ptak, 1999). Failure in the second run is seen in patients with anterior limbic lesions, in particular of
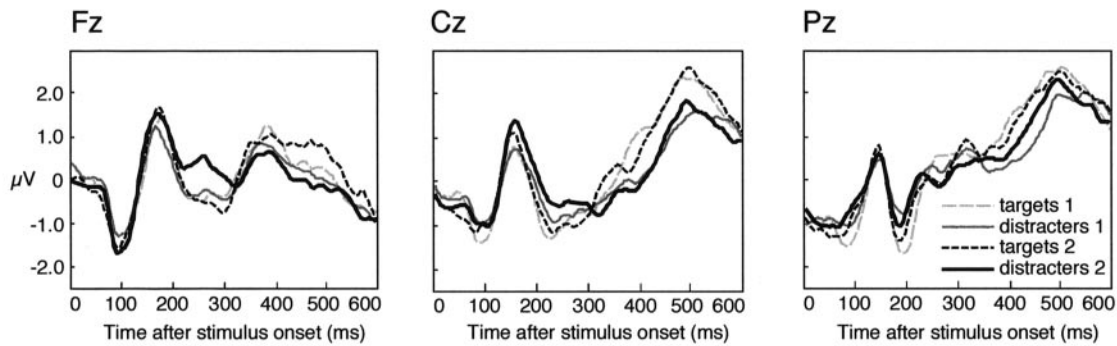
Figure 2. Evoked potential curves. Responses over frontal electrode Fz, central Cz and posterior parietal Pz for the four stimulus types: targets and distracters in run 1 and run 2.

the OFC or its connections (Schnider *et al.*, 1996b,c, 2000a; Schnider and Ptak, 1999). PET showed that healthy subjects performing this run had circumscribed activation of the posterior medial OFC (Schnider *et al.*, 2000b).

ERP Analysis
The electroencephalogram (EEG) was continuously recorded from 128 electrodes (Electrical Geodesics, Inc., OR, USA) at a sampling rate of 250 Hz and a bandpass of 0.1–100 Hz. Electrode impedance was kept below 50 KΩ. Electrodes were referenced to Cz and recomputed off-line to the average reference. Artifact-free EEG responses were aligned off-line to stimulus onset and averaged (ERPs). The first 600 ms of the ERPs after stimulus onset were retained for analysis. Conventional analysis on electrodes of interest and global spatial analysis were performed.

Waveform Analysis
Conventional waveform analysis of the electrical responses to the four stimulus types was performed over the electrode positions commonly considered in ERP studies, i.e. at frontal electrode Fz, vertex electrode Cz and the posterior parietal electrode Pz (against the average reference). Apparent amplitude differences between the ERP traces in response to the four stimulus types were tested for statistical significance using repeated measures ANOVAs with the two factors Stimulus type (target, distracter) and Run (run 1, run 2).

Global Spatiotemporal Analysis
The global spatiotemporal analysis consisted in the segmentation of the ERPs into successive time periods of stable scalp electric field configurations (stable map topographies). Transiently stable potential configurations have been observed with many cognitive tasks and are assumed to reflect the basic functional microstates of the brain during information processing (Lehmann and Skrandies, 1980; Lehmann, 1987; Michel *et al.*, 1992, 1999; Pegna *et al.*, 1997; Khateb *et al.*, 2000; Morand *et al.*, 2000; Thut *et al.*, 2000). The spatiotemporal segmentation procedures allow us to determine the sequence of these microstates and their specificity in a given condition. The segmentation procedure used here has been described in detail elsewhere (Khateb *et al.*, 1999). In brief, the Global Dissimilarity (an index of configuration difference between successive maps independent of strength) (Lehmann, 1987) was calculated for each grand-mean ERP of each of the four stimulus types (distracters and target of runs 1 and 2). Maps between dissimilarity peaks were considered similar and averaged. Using a spatial correlation matrix, the remaining maps were further reduced to the most significant field configurations appearing in all four conditions. The appearance of these maps in the grand mean and in the individual data was then determined by a fitting procedure that allowed us to determine how well the maps explained the individually obtained patterns of activity (global explained variance, GEV), their duration, and the time point at which they best explain the individual data (time point of best fit) (Pascual-Marqui *et al.*, 1995; Pegna *et al.*, 1997; Michel *et al.*, 1999; Morand *et al.*, 2000). These individual measures (GEV of the maps and duration during which they were present) were then subjected to repeated-measures ANOVAs with the two factors Stimulus type (target, distracter) and Run (run 1, run 2).
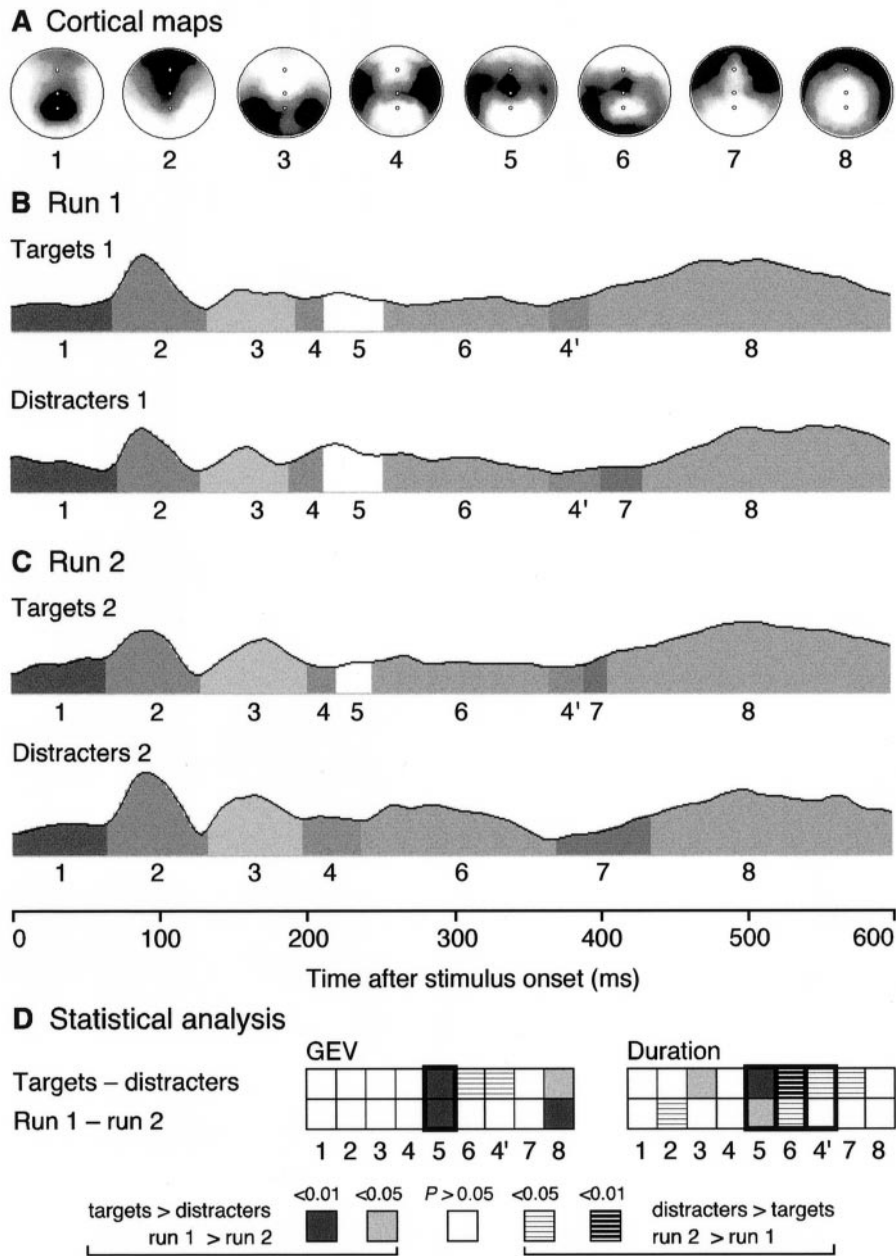
Results

Behavioral Results
The performance of our 13 test subjects in this task was almost perfect: hit rate (maximum 72) was 66.1 ± 3.6 in the first run and 63.1 ± 6.6 in the second run [repeated-measures ANOVA: $F(1,12) = 2.5$; $P = 0.14$]; false positive rate (maximum 96) was minimal, but significantly increased from the first (1.7 ± 1.7) to the second run (2.9 ± 1.9) [$F(1,12) = 6,2$; $P = 0.03$]. The small number of false responses precluded an analysis of the brain activity related to errors. Reaction times varied between the stimulus types [$F(3,13) = 6$; $P = 0.002$]; they were fastest in response to distracters of run 1 (715 ± 96 ms) and slowest to targets of run 2 (790 ± 112 ms).

Conventional ERP Analysis
Figure 2 displays the ERPs obtained in response to the four stimulus types at electrode positions Fz, vertex electrode Cz and the posterior parietal electrode Pz. Two periods with distinct and consistent amplitude differences appear evident. Between about 200 and 300 ms after stimulus presentation, the amplitude of the ERP in response to distracters of run 2 is clearly distinct from the other three stimulus types at electrode Fz, less so over Cz. In contrast to the other stimulus types, which have a negative deflection in this period, distracters of run 2 do not induce such a negativity; the potential remains positive. A two-factor repeated-measures ANOVA using each subject's mean potential amplitude over Fz in the interval between 200 and 300 ms revealed a significant effect of Stimulus type [$F(1,12) = 5.3$; $P = 0.04$] and interaction of Stimulus type × Run [$F(1,12) = 50$; $P = 0.04$], which was solely due to the higher amplitude of the response to the distracters of run 2. Over Cz, the result was essentially similar [effect of Stimulus type: $F(1,12) = 18$; $P = 0.001$] with a more positive amplitude in response to distracters than targets only in run 2, but the interaction fell short of significance [$F(1,12) = 3.7$; $P = 0.08$].

A second period with apparently different ERPs emerges at about 400–500 ms, when the responses to targets appear to have a higher (more positive) amplitude than responses to distracters. This difference appears most pronounced over Cz and Pz for the stimuli of run 1, over Fz for the stimuli of run 2. An ANOVA comparing the mean potential amplitudes in the interval between 400 and 500 ms confirmed this impression: at Fz, there was a significant main effect of Stimulus type [$F(1,12) = 8.5$; $P = 0.01$], which was due to a significant amplitude difference

Figure 3. Results of segmentation. (A) Eight cortical maps obtained from segmentation of the grand mean of all subjects within the first 600 ms. White indicates positivity, black indicates negativity in reference to the average potential amplitude. The electrode positions of the waveform analysis (Fig. 2) are indicated as fine dots in the maps. (B,C) Distribution of stable map configurations (segments) over 600 ms after stimulus onset in run 1 (B) and run 2 (C). Each segment has a separate shade; the numbers below each row indicate the map, as shown in (A), representative of this period. 4′ denotes the second appearance of map 4. The amplitude of the curves in (B) and (C) reflects the global field power. (D) Result of *post hoc* tests of the repeated measures ANOVAs with the two factors Stimulus type (target, distracter) and Run (run 1, run 2). The rectangles with thick lines indicate maps with a significant interaction of Stimulus type × Run.

between targets and distracters in run 2, but not run 1. Over Cz, there was a significant effect of Stimulus type [$F(1,12)$ = 15.3; P = 0.002] due to targets inducing more positive amplitudes than distracters in both runs. Over Pz, there was also a significant effect of Stimulus type [$F(1,12)$ = 22; P = 0.0005] with targets having higher amplitudes than distracters. In addition, there was a significant interaction of Stimulus type × Run [$F(1,12)$ = 11.8; P = 0.005], which was explained by the fact that only in run 1

did distracters (very first presentation of the stimuli) induce a significantly lower (more negative) potential than targets.

In summary, in both runs, distracters differ from targets after ~400–500 ms with differences being more pronounced over the two posterior electrodes in run 1 and over the two anterior electrodes in run 2. The most striking finding, however, is a distinctly different ERP response to distracters of run 2 over frontal electrode Fz at ~200–300 ms.

Potential Map Configurations

Event-related potentials evoked by correct target and distracter detection in the first and second run were analyzed separately. Temporal segmentation applied to the grand mean of all test subjects yielded eight segments with distinct cortical maps within the first 600 ms after stimulus presentation (Fig. 3A). As suggested in Figure 3B,C, there are two time periods with different map configurations between the four stimulus conditions: at ~220 ms, map 5 does not appear in response to the distracters of run 2 (Fig. 3C, 'distracters'). At ~400 ms, maps 4' and 7 appear to differ among the stimulus types (Fig. 3B,C). This initial inspection of the maps thus indicates differences in the same periods, in which the waveform analysis (which looked at only 3 of the 128 electrodes constituting the map) had demonstrated different stimulus responses. Most importantly, it appears that the frontal positivity in response to distracters of run 2 at 200–300 ms (Fig. 2, Fz) is associated with the absence of a particular map configuration (map 5), rather than a distinct map configuration which would have indicated an additional processing stage specific to the treatment of distracters in run 2.

The statistical analysis on the results of the map fitting procedure in the individual data is summarized in Figure 3D. This analysis confirmed the absence of map 5 in response to the distracters of run 2. The map, which had a time point of best fit at 237 ± 27 ms (onset 217 ± 23 ms; offset 273 ± 25 ms), showed significant main effects regarding its GEV (weaker in distracters and in run 2) and duration (shorter in response to distracters and in run 2) (Fig. 3D). Most importantly, there was a significant interaction of Stimulus type × Run regarding the GEV [$F(1,12) = 4.9$; $P = 0.047$] and the duration [$F(1,12) = 5.2$; $P = 0.042$]. These interactions were explained by a significantly lower GEV and shorter duration of map 5 in response to the distracters of run 2. Probably in compensation of this shortening of map 5, map 6 was significantly longer in response to distracters of run 2 [significant main effects of Stimulus type and Run (Fig. 3D); and interaction of Stimulus type × Run, $F(1,12) = 5.2$; $P = 0.042$].

Statistical analysis also confirmed the differences at ~400 ms (Fig. 3B,C). Map 4' (time point of best fit 392 ± 67 ms; onset 353 ± 55 ms; offset 422 ± 79 ms) had a stronger GEV and was longer in response to distracters (Fig. 2D). This result appears surprising given the absence of map 4' in response to distracters of run 2 from the grand mean displayed in Figure 3C. However, there was a significant interaction of Stimulus type × Run regarding the duration of map 4' ($F(1,12) = 5.4$; $P = 0.038$) which was due to a significantly longer duration of map 4' in response to distracters than targets in run 1, when the stimuli are presented for the very first time, but not in run 2. Map 7 (time point of best fit 456 ± 86 ms; onset 392 ± 75 ms; offset 517 ± 82 ms) also had a longer duration in response to distracters, but no significant interaction. Thus, the cortical maps (GEV and duration) differ after ~400–500 ms between distracters and targets in general, but less consistently between the runs. They thus most likely reflect processes of learning (consolidation) and recognition.

Discussion

This study shows that when subjects successfully distinguish between memories that pertain to ongoing reality and memories that do not, there is a specific alteration of cortical electrical activity 220–300 ms after presentation of items that were previously seen (in the first run) but have no current relevance (distracters of run 2). In comparison with the processing of new information (distracters of run 1) and of currently pertinent memories (defined as stimulus repetitions within the run, i.e.

targets of both runs), the stimuli evoking a currently irrelevant memory (distracters of run 2) specifically failed to induce the negative deflection of a frontal potential after 200–300 ms which was induced by the other stimuli (Fig. 2). In terms of cortical network activity, as reflected in electrical cortical map configurations, the processing of these 'currently irrelevant' memories (distracters of run 2) induced absence of a particular cortical potential map configuration after 220–300 ms (absence of map 5, Fig. 3C), a finding indicating that these stimuli induce suppression of a processing stage, rather than an additional cortical process. The finding that among the four stimulus types appearing in our task (distracters and targets in run 1 and 2), there is a specific alteration of cortical activity exclusively in response to distracters of run 2, is in agreement with our previous clinical studies. Patients having anterior limbic lesions specifically failed to suppress the distracters of the second run during the period of their disease in which they confused old memories with ongoing reality (Schnider and Ptak, 1999; Schnider et al., 2000a). Together, these results support the idea that the brain sorts out memories that pertain to ongoing reality by suppressing currently irrelevant memories (Schnider and Ptak, 1999; Schnider et al., 2000a).

The concurrence of clinical (Schnider et al., 1996b, 2000a; Schnider and Ptak, 1999), imaging (Schnider et al., 2000b) and electrophysiological results (present study) underscores the validity of our experimental paradigm as a tool to study the brain's monitoring of ongoing reality, i.e. the selection of memories that pertain to current reality. It may be of relevance that the task used in this study was considerably more difficult than the two task variations used with patients (Schnider et al., 1996b,c; Ptak and Schnider, 1999; Schnider and Ptak, 1999; Ptak et al., 2001): rather than presenting a few target stimuli several times during a run, target stimuli continuously changed during the two runs of the present experiment. This adaptation was made in order to force subjects to remain attentive throughout the task, and because healthy subjects performing the easier clinical version of the task had indicated that they had became aware of the number and content of target items. Most importantly, difficulty of the present task was increased by making the two runs in rapid succession, rather than separated by 1 h. This change was made for two reasons. Firstly, both healthy subjects and non-confabulating amnesics had had some difficulty in suppressing the interference by distracters of the second run only when the runs were made in immediate succession, but not when they were separated by 5 min (Schnider and Ptak, 1999); in comparison, spontaneously confabulating patients had failed to suppress this interference even when the runs were separated by 30 min (Schnider and Ptak, 1999) or 1 h (Schnider et al., 1996b). Secondly, our previous PET study had shown that healthy subjects conceived of the task as challenging and that OFC activation was obtained only when the runs were made in rapid succession (Schnider et al., 2000b). The observation that the task needed to be adapted in healthy subjects in order to produce the difficulty experienced by spontaneous confabulators in the easier task, indicates that whereas failure of this orbitofrontal suppression capacity is associated with intrusion of memories over long periods — 1 h in the clinical version of our task, sometimes years in the patients' everyday life, as evident from spontaneous confabulation — this process appears to be highly efficient and rapid in the healthy brain, possibly even continuously updating our interpretation of ongoing reality.

It is not immediately obvious why, in comparison with traditional temporal order and recency tasks, the second run of our task has such strong behavioral and anatomical specificity.

Traditional tasks probe conscious knowledge about when and where information was last encountered. Failure in these tasks may occur together with amnesia, but also as an isolated finding in patients having neither amnesia nor confabulations (Huppert and Piercy, 1976; Shimamura *et al.*, 1990; Milner *et al.*, 1991; Johnson *et al.*, 1997b; Kopelman *et al.*, 1997b). The isolated failure in these tasks is typically seen after dorsolateral prefrontal lesions (Shimamura *et al.*, 1990; Milner *et al.*, 1991; Kopelman *et al.*, 1997b), a lesion site which in our studies was sometimes associated with failure in the first run of our task (learning), but never with suppression failure in the second run or with spontaneous confabulation (Schnider *et al.*, 1996b,c; Schnider and Ptak, 1999). Failure in traditional source and temporal order memory in the context of an amnesia has no anatomical specificity beyond the amnesia itself (Bowers *et al.*, 1988; Janowsky *et al.*, 1989; Shimamura *et al.*, 1990; Milner *et al.*, 1991; Parkin and Hunkin, 1993; Kesner *et al.*, 1994; Schnider *et al.*, 1996a; Johnson *et al.*, 1997b; Kopelman *et al.*, 1997b). The same holds true for the intrusions (often from prior lists or previous tasks requiring word production) that some memory-impaired subjects make when recalling a learned list of words. Even though these intrusions also constitute a form of 'currently irrelevant memories' and may in that sense resemble the memory intrusions determining the interpretation of ongoing reality in a spontaneous confabulator's thinking, the two phenomena (spontaneous confabulation and intrusions) are entirely independent: firstly, they doubly dissociate, i.e. the occurrence of spontaneous confabulation is independent of the production of intrusions, and vice versa. Secondly, the task used in the present study, which so perfectly predicts spontaneous confabulation and orientation, has no predictive value (no significant correlation) for intrusions in a memory test (Schnider *et al.*, 1996b).

Imaging studies with healthy subjects underscore the specificity of our task: correct performance in classical source and temporal order tasks induced lateral prefrontal activation (Zorrilla *et al.*, 1996; Cabeza *et al.*, 1997; Rugg *et al.*, 1999), rather than the orbitofrontal activation associated with our task (Schnider *et al.*, 2000b). Similarly, tasks manipulating encoding or retrieval conditions or specifically demanding knowledge about whether information was previously presented in a particular context or modality [the 'exclusion' condition of Jacoby (Jacoby, 1991)] consistently activated lateral frontal cortex, never the orbitofrontal cortex activated by our task [as reviewed elsewhere (Fletcher and Henson, 2001)]. The same is true for tasks requiring distinction between items that were truly presented before, and similarly appearing items that were not really presented (illusory memories); again, activation of the lateral frontal (plus other brain areas, e.g. medial temporal), rather than medial orbital frontal cortex was observed (Schacter *et al.*, 1996, 1997). It is important to realize that these memory tasks demand monitoring of whether an item was really presented in a particular situation in the past (Johnson *et al.*, 1993; Johnson and Raye, 1998; Schacter *et al.*, 1998). This is not required by our task; all items have previously been presented in the same setting. The second run demands the ability to separate the memory of a previous episode (the first run) from memories put on hold during the episode constituting current reality – the ongoing second run. Familiarity judgements, which helped to perform the first run (a familiar item could be assumed to be a target), are no more helpful. The second run of our task may provide such a powerful paradigm for exploring the monitoring of ongoing reality because – unlike common recognition, source monitoring or 'exclusion' tasks – it is does not differ from the first run in terms of information content, context or source.

These observations also suggest that correct interpretation of ongoing reality, i.e. the continued activity of memory traces that pertain to it, depends on an on-line filtering mechanism matching currently active memory traces with currently available external information, rather than on construction of a reality representation based on the recency of activated memory traces.

Our study demonstrates that that suppression of interference by memories that do not pertain to ongoing reality (processing the distracters of run 2) is reflected in distinct alteration of cortical activity at 220–300 ms. Compared with this process, learning and recognition appear to be cortically expressed at a relatively late stage of processing, after ~400–480 ms. This conclusion is based on the finding that the electrical response to the items' first presentation and their recurrence, particularly during the first (learning) run, is markedly different in this period. The finding agrees with earlier studies demonstrating a stronger potential amplitude in this late period after presentation of previously seen than new items (Friedman, 1990a,b; Bentin *et al.*, 1992; Retzlaff and Morris, 1996). Our study demonstrates that learning and recognition are cortically distinguished by varying potential amplitudes (Fig. 2) rather than a distinct pattern of electrical cortical activity (Fig. 3). That is, learning and recognition appear to depend on the modulation of cortical networks that are common to the processing both of novel and familiar information. This finding is in agreement with models of cortical memory which postulate that perception and memory rely on the activity of similar neuronal networks (Damasio, 1989; Fuster, 1995; Singer, 1998).

These results show that before memories enter the stage of learning and recognition, their cortical representation is specifically adapted according to whether they relate to ongoing reality or not. This finding has several implications. First, assuming that the modulation of cortical activity after 400–480 ms constitutes an essential processing stage for conscious recollection of memories, it suggests that the suppression of memories that do not pertain to ongoing reality is a pre-conscious process. At any rate, it occurs much earlier than processes presumed to monitor the veridicality and source of evoked memories (Burgess and Shallice, 1996; Johnson and Raye, 1998; Schacter *et al.*, 1998; Melo *et al.*, 1999). Indeed, cortical activity associated with source monitoring and distinction between true and illusory memories appears even later than the 400–480 ms associated with encoding and recognition (Johnson *et al.*, 1997a). If suppression of information that does not pertain to ongoing reality is indeed a pre-conscious process, it easily explains that patients with anterior limbic lesions, who fail in this suppression capacity (spontaneous confabulators), are absolutely convinced about the veracity of their false memories and often act on their basis (Korsakow, 1892; Van der Horst, 1932; Dalla Barba, 1993; Schnider *et al.*, 1996a,b,c; Kopelman *et al.*, 1997a; Burgess and McNeil, 1999; Ptak and Schnider, 1999; Schnider and Ptak, 1999). Secondly, given that memories are monitored for their relation with current reality and pre-processed before they are again consolidated, it is possible that this mechanism also contributes to our feeling for what memory refers to a true event and what memory refers to a fantasy. The early adjustment of cortical activity, depending on whether the processed information relates to ongoing reality or not, may prepare information to be stored as a true event or as a thought. This mechanism might thus be one prerequisite for post-retrieval processes monitoring the veridicality of evoked memories to be successful (Burgess and Shallice, 1996; Johnson and Raye, 1998; Schacter *et al.*, 1998; Melo *et al.*, 1999).

The cortical distinction between memories that pertain to

ongoing reality and memories that do not, which occurs after 220–300 ms, is an early process in relation to the recognition and consolidation of memories (400–480 ms), but is not particularly early in relation to many other cognitive processes associated with distinct cortical potentials at this stage. Based on our previous observations that failure in the second run of our task occurs exclusively in patients having anterior limbic damage (Schnider *et al.*, 1996c; Schnider and Ptak, 1999), and that correct performance in the second run activates the posterior medial OFC of healthy subjects (Schnider *et al.*, 2000b), we presume that the alteration of cortical activity after 220–300 ms is but the cortical expression of an anterior limbic, in particular posterior orbitofrontal, suppression mechanism that actually occurs much earlier. Testing of this theory would require recording of electrical activity directly from the OFC.

The role of posterior medial OFC in the ability to adjust thinking to ongoing reality and to prevent behavior from being based on memories that have no current relevance is also supported by animal studies which have shown a comparable failure after lesions similar to the lesions of spontaneous confabulators. Monkeys with ablations of the posterior medial OFC (but not of the lateral or anterior OFC or the dorsolateral prefrontal cortex) had a severe deficit of extinction, i.e. they continued to react to stimuli which were no longer rewarded (Butter, 1969). In other words, following omission of an expected reward, monkeys with posterior medial OFC ablation fail to adapt their behavior and continue to act on the basis of the now irrelevant memory that the cue was previously followed by reward. The posterior medial OFC of monkeys contains neurons that specifically increase their firing rate when an expected reward fails to be delivered, i.e. they specifically fire in extinction trials (Rosenkilde *et al.*, 1981; Thorpe *et al.*, 1983). Thus, these neurons signal the inconsistency of a memory with ongoing reality (absence of the reward expected to follow a cue). If one accepts the idea that human behavior, too, is motivated by predicted goals, this model may be applied to the human ability to adapt behavior and thinking to a changing reality.

How might neurons in the posterior medial OFC suppress the influence of currently irrelevant memories on thinking and behavior? The physiological mechanism of the suppression process, whose failure is associated with spontaneous confabulation and disorientation (Schnider *et al.*, 1996b,c, 2000a; Schnider and Ptak, 1999), is obviously unknown. Assuming that activated memories are cortically represented as the synchronous activity of neuronal populations (Fuster, 1995; Singer, 1998), suppression of currently irrelevant memories might be exerted by simple desynchronization of these populations (Singer, 1998). This idea would be in accord with our present finding that processing of distracters in run 2 was accompanied by the absence of a specific potential map configuration, which was present when the other types of stimuli were processed (map 5, Fig. 3*B*,*C*). But how, and through what anatomical routes, might the OFC induce desynchronization of cortical neuronal networks; how does it communicate with the neocortex? One suggestion has been that frontal subcortical loops, which connect frontal cortex with distinct portions of the striatum, pallidum and substantia nigra, thalamic nuclei, and projections back to the cortex (Alexander *et al.*, 1986), might allow communication between the OFC and the neocortex (Rolls, 1999). Whereas the loops emanating from different cortical areas appear to be fairly segregated down to the level of the striatum (Haber *et al.*, 1995), immense cross-communication and convergence of fibers seem to be present at the level of the substantia nigra (Percheron *et al.*, 1994). The orbitofrontal

cortex, which initially projects to the ventral striatum (Haber *et al.*, 1995), might thus influence activity of wide areas of the neocortex. It is noteworthy that the non-delivery of an expected reward, which elicits increased firing of selected neurons in the OFC, has been shown to inhibit firing of select neurons in the ventral striatum and the substantia nigra at the time when the reward was expected to be delivered (Schultz *et al.*, 1992). This finding would be compatible with the idea that the OFC may actively suppress neuronal activity in subsequent sites of the loop connecting it with the neocortex.

The cited studies let us suggest a theory on how the brain monitors ongoing reality. It appears that the anterior limbic system, in particular the posterior medial OFC, is the essential mediator of the ability to refer thinking and behavior to ongoing reality (Schnider *et al.*, 1996b,c, 2000a). It exerts this role by constantly suppressing activated memory traces that do not relate to ongoing reality (Schnider and Ptak, 1999; Schnider *et al.*, 2000b). This suppression influences the cortical representation of memories before their content is recognized and consolidated (present study). Findings from animal studies would be compatible with the idea that human reality monitoring reflects a distinct property of the brain's reward system.

Notes

References

Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci 9:357–381.

Bentin S, Moscovitch M, Heth I (1992) Memory with and without awareness: performance and electrophysiological evidence of savings. J Exp Psychol Learn Mem Cogn 18:1270–1283.

Bowers D, Verfallie M, Valenstein E, Heilman KM (1988) Impaired acquisition of temporal information in retrosplenial amnesia. Brain Cogn 8: 47–66.

Burgess PW, McNeil JE (1999) Content-specific confabulation. Cortex 35:163–182.

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

Butter CM (1969) Perseveration in extinction and in discrimination reversal tasks following selective frontal ablations in *Macaca mulatta*. Physiol Behav 4:163–171.

Cabeza R, Mangels J, Nyberg L, Habib R, Houle S, McIntosh AR, *et al.* (1997) Brain regions differentially involved in remembering what and when: a PET study. Neuron 19:863–870.

Dalla Barba G (1993) Confabulation: knowledge and recollective experience. Cogn Neuropsychol 10:1–20.

Damasio AR (1989) Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. Cognition 33:25–62.

DeLuca J, Cicerone kDa (1991) Confabulation following aneurysm of the anterior communicating artery. Cortex 27:417–423.

Fletcher PC, Henson RN (2001) Frontal lobes and human memory: insights from functional neuroimaging. Brain 124: 849–881.

Friedman D (1990a) Cognitive event-related potential components during continuous recognition memory for pictures. Psychophysiology 27:136–148.

Friedman D (1990b) ERPs during continuous recognition memory for words. Biol Psychol 30:61–87.

Fuster JM (1995) Memory in the cerebral cortex. Cambridge, MA: MIT Press.

Haber SN, Kunishio K, Mizobuchi M, Lynd-Balta E (1995) The orbital and medial prefrontal circuit through the primate basal ganglia. J Neurosci 15: 4851–4867.

Huppert FA, Piercy M (1976) Recognition memory in amnesic patients: effect of temporal context and familiarity of material. Cortex 12:3–20.

Jacoby LL (1991) A process dissociation framework: separating automatic from intentional uses of memory. J Mem Lang 30:513–541.

Janowsky JS, Shimamura AP, Squire LR (1989) Source memory impairment in patients with frontal lobe lesions. Neuropsychologia 27:1043–56.

Johnson MK, Raye CL (1998) False memories and confabulation. Trends Cogn Sci 2:137–145.

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

Johnson MK, Kounios J, Nolde SF (1997a) Electrophysiological brain activity and memory source monitoring. NeuroReport 8:1317–1320.

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

Kesner RP, Hopkins RO, Fineman B (1994) Item and order dissociation in humans with prefrontal damage. Neuropsychologia 32:881–889.

Khateb A, Annoni JM, Landis T, Pegna AJ, Custodi MC, Fonteneau E, *et al.* (1999) Spatio-temporal analysis of electric brain activity during semantic and phonological word processing. Int J Psychophysiol 32:215–231.

Khateb A, Michel CM, Pegna AJ, Landis T, Annoni JM (2000) New insights into the Stroop effect: a spatio-temporal analysis of electric brain activity. NeuroReport 11:1849–1855.

Kopelman MD, Hg N, Van den Brouke O (1997a) Confabulation extending over episodic, personal, and general semantic memory. Cogn Neuropsychol 14:683–712.

Kopelman MD, Stanhope N, Kingsley D (1997b) Temporal and spatial context memory in patients with focal frontal, temporal lobe, and diencephalic lesions. Neuropsychologia 35:1533–1545.

Korsakow SS (1892) Erinnerungstäuschungen (Pseudoreminiscenzen) bei polyneuritischer Psychose. Allg Z Psychiat Psych Med 47:390–410.

Lehmann D (1987) Principles of spatial analysis. In: Handbook of electroencephalography and clinical neurophysiology, Vol. 1: Methods of analysis of brain electrical and magnetic signals (Gevins AS and Rémond A, eds), pp. 309–354. Amsterdam: Elsevier.

Lehmann D, Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroenceph Clin Neurophysiol 48:609–621.

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. Cogn Neuropsychol 16:343–359.

Michel CM, Henggeler B, Lehmann D (1992) 42-channel potential map series to visual contrast and stereo stimuli: perceptual and cognitive event-related segments. Int J Psychophysiol 12:133–145.

Michel CM, Seeck M, Landis T (1999) Spatio-temporal dynamics of human cognition. NIPS 14:206–214.

Milner B, Corsi P, Leonard G (1991) Frontal-lobe contribution to recency judgements. Neuropsychologia 29:601–618.

Morand S, Thut G, de Peralta RG, Clarke S, Khateb A, Landis T, *et al.* (2000) Electrophysiological evidence for fast visual processing through the human koniocellular pathway when stimuli move. Cereb Cortex 10:817–825.

Parkin AJ, Hunkin NM (1993) Impaired temporal context memory on anterograde but not retrograde tests in the absence of frontal pathology. Cortex 29:267–280.

Pascual-Marqui RD, Michel CM, Lehmann D (1995) Segmentation of brain electrical activity into microstates: model estimation and validation. IEEE Trans Biomed Engng 42:658–665.

Pegna AJ, Khateb A, Spinelli L, Seeck M, Landis T, Michel CM (1997) Unraveling the cerebral dynamics of mental imagery. Hum Brain Mapp 5:410–421.

Percheron G, Yelnik J, Francois C, Fenelon G, Talbi B (1994) Analyse informationnelle du système lié aux ganglions de la base. Rev Neurol 150: 614–626.

Ptak R, Schnider A (1999) Spontaneous confabulations after orbitofrontal damage: the role of temporal context confusion and self-monitoring. Neurocase 5:243–250.

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

Retzlaff PD, Morris GL (1996) Event-related potentials during the Continuous Visual Memory Test. J Clin Psychol 52:43–47.

Rolls ET (1999) The brain and emotion. Oxford: Oxford University Press.

Rosenkilde CE, Bauer RH, Fuster JM (1981) Single cell activity in ventral prefrontal cortex of behaving monkeys. Brain Res 209:375–394.

Rugg MD, Fletcher PC, Chua PM, Dolan RJ (1999) The role of the prefrontal cortex in recognition memory and memory for source: an fMRI Study. NeuroImage 10:520–529.

Schacter DL, Reiman E, Curran T, Yun LS, Bandy D, McDermott KB, *et al.* (1996) Neuroanatomical correlates of veridical and illusory recognition memory: evidence from positron emission tomography. Neuron 17:267–274.

Schacter DL, Buckner RL, Koutstaal W, Dale AM, Rosen BR (1997) Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. NeuroImage 6: 259–269.

Schacter DL, Norman KA, Koutstaal W (1998) The cognitive neuroscience of constructive memory. Annu Rev Psychol 49:289–318.

Schnider A, Ptak R (1999) Spontaneous confabulators fail to suppress currently irrelevant memory traces. Nat Neurosci 2:677–681.

Schnider A, Gutbrod K, Hess CW, Schroth G (1996a) Memory without context. Amnesia with confabulations following right capsular genu infarction. J Neurol Neurosurg Psychiat 61:186–193.

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

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

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

Schnider A, Treyer V, Buck A (2000b) Selection of currently relevant memories by the human posterior medial orbitofrontal cortex. J Neurosci 20:5880–5884.

Schultz W, Apicella P, Scarnati E, Ljungberg T (1992) Neuronal activity in monkey ventral striatum related to the expectation of reward. J Neurosci 12:4595–4610.

Shimamura AP, Janowsky JS, Squire LR (1990) Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. Neuropsychologia 28:803–813.

Singer W (1998) Consciousness and the structure of neuronal representations. Philos Trans R Soc Lond B Biol Sci 353:1829–1840.

Snodgrass JG, Vanderwart M (1980) A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol Hum Learn Mem 6:174–215.

Thorpe SJ, Rolls ET, Maddison S (1983) The orbitofrontal cortex: neuronal activity in the behaving monkey. Exp Brain Res 49:93–115.

Thut G, Hauert C, Viviani P, Morand S, Spinelli L, Blanke O, *et al.* (2000) Internally driven vs. externally cued movement selection: a study on the timing of brain activity. Brain Res Cogn Brain Res 9:261–269.

Van der Horst L (1932) Über die Psychologie des Korsakowsyndroms. Monatsschr Psychiat Neurol 83:65–84.

Zorrilla LT, Aguirre GK, Zarahn E, Cannon TD, D'Esposito M (1996) Activation of the prefrontal cortex during judgments of recency: a functional MRI study. NeuroReport 7: 2803–2806.