

Neural Precursors of Delayed Insight

Annabelle Darsaud¹, Ullrich Wagner², Evelyne Balteau¹,
Martin Desseilles¹, Virginie Sterpenich¹, Gilles Vandewalle¹,
Geneviève Albouy¹, Thanh Dang-Vu¹, Fabienne Collette¹,
Melanie Boly¹, Manuel Schabus¹, Christian Degueldre¹,
Andre Luxen¹, and Pierre Maquet¹

Abstract

■ The solution of a problem left unresolved in the evening can sometimes pop into mind as a sudden insight after a night of sleep in the following morning. Although favorable effects of sleep on insightful behavior have been experimentally confirmed, the neural mechanisms determining this delayed insight remain unknown. Here, using fMRI, we characterize the neural precursors of delayed insight in the number reduction task (NRT), in which a hidden task structure can be learned implicitly, but can also be recognized explicitly in an insightful process, allowing immediate qualitative improvement in task performance. Normal volunteers practiced the NRT during two fMRI sessions (training and retest), taking place 12 hours apart after a night of sleep. After this delay, half of the subjects gained insight into the hidden task structure (“solvers,” S), whereas the other half did not (“nonsolvers,” NS). Already at training, solvers and nonsolvers differed in their cerebral responses associated with implicit learning. In future solvers, responses were observed in the superior frontal sulcus, posterior

parietal cortex, and the insula, three areas mediating controlled processes and supporting early learning and novice performance. In contrast, implicit learning was related to significant responses in the hippocampus in nonsolvers. Moreover, the hippocampus was functionally coupled with the basal ganglia in nonsolvers and with the superior frontal sulcus in solvers, thus potentially biasing participants’ strategy towards implicit or controlled processes of memory encoding, respectively. Furthermore, in solvers but not in nonsolvers, response patterns were further transformed overnight, with enhanced responses in ventral medial prefrontal cortex, an area previously implicated in the consolidation of declarative memory. During retest in solvers, before they gain insight into the hidden rule, significant responses were observed in the same medial prefrontal area. After insight, a distributed set of parietal and frontal areas is recruited among which information concerning the hidden rule can be shared in a so-called global workspace. ■

INTRODUCTION

Insight refers to the penetrating and often sudden conscious understanding of the covert structure of a problem, which results in the conception of its optimal solution. The experimental characterization of this elusive phenomenon took advantage, among other experimental strategies, of implicit learning paradigms in which participants suddenly gain insight in a hidden task regularity, originally unknown to them (Haider & Rose, 2007). For instance, in the number reduction task (NRT), participants have to transform a given eight-digit sequence into a sequence of seven responses to determine a specific digit as the final result of the trial. This can be achieved by successive application of two simple, explicitly known, transformation rules to pairs of digits. Unknown to participants, however, a hidden structure is implemented in all trials that, when discerned, allows an early determination of the solution digit: The last three responses mirror the three previous ones such that

the second response digit always coincides with the solution. When participants gain insight into the hidden structure, they cut short sequential responding, anticipatively confirming the second response digit as the final result. The sudden drop in the time needed to reach the solution of a given trial identifies the moment at which the insight occurs (Haider & Rose, 2007).

This paradigm is remarkable because it allows investigating how an explicit knowledge emerges from an implicit premise. Classically, human memory involves distinct systems which can be behaviorally, functionally, and anatomically dissociated (Schacter & Tulving, 1994). Declarative memory refers to the capacity to explicitly and verbally recall facts and events (Squire & Zola, 1996). It crucially involves the hippocampus and surrounding medial temporal cortices (Eichenbaum, 2004). In contrast, procedural memory is progressively acquired through repeated practice and manifests itself by individual skills and habits that cannot be easily verbalized or otherwise be explicitly specified. In the motor domain, memory representations mainly involve cerebello-cortical and striato-cortical networks

¹University of Liège, Belgium, ²Bangor University, United Kingdom

(Doyon & Benali, 2005). At odds with this classical view, an increasing body of evidence indicates that both hippocampal and striatal systems can be recruited during procedural learning and can even interact to support learned behavior (Poldrack & Rodriguez, 2004; Poldrack & Packard, 2003; Poldrack et al., 2001). For instance, the hippocampus is involved during motor sequence learning, even when participants remain unaware of the sequential succession of learned movements (Albouy et al., 2008; Schendan, Searl, Melrose, & Stern, 2003). Accordingly, for the NRT, implicitly learning the hidden structure recruits perirhinal cortex, whereas responses in the basal ganglia and cerebellum merely increase with practice (Rose, Haider, Weiller, & Buchel, 2002). Using other experimental paradigms than the NRT, the insight itself was associated with responses in the hippocampus (Luo & Niki, 2003) and various cortical areas such as the precuneus and medial frontal areas (Jung-Beeman et al., 2004; Luo & Niki, 2003).

Based on these findings, two hypotheses can account for the emergence of insight. One assumes that declarative and procedural knowledge are encoded in distinct independent memory systems. In this case, gaining insight into the hidden structure of the NRT would correspond to an irreversible bias of participants' strategy toward the explicit representation. An alternative hypothesis would consider that the functional interactions between hippocampus-dependent and striato-cerebello-cortical memory systems constitute the central phenomenon that leads to the emergence of insight. In this case, implicit processes taking place during initial practice might provide the primitive memory elements on which insight eventually builds up.

Another important aspect of insightful behavior is that insight does not necessarily happen during the initial exposure to the intractable problem, but can emerge at a later date, especially after sleep (Wagner, Gais, Haider, Verleger, & Born, 2004). This time course suggests that, in this case, the task representations generated during the initial exposure have to be processed off-line, that is, without additional practice, to form novel associations and subsequently lead to a qualitative modification in behavior. This delayed effect also suggests that the initial memory representations generated during training differ between participants that will subsequently gain insight into the hidden rule ("solvers," S) and those who will not ("non-solvers," NS) (Maquet & Ruby, 2004).

Here, we used fMRI to assess brain responses elicited by learning the NRT during two separate sessions (training and retest), 12 hours apart. In order to favor insight and obtain comparable proportions of solver and nonsolver participants during retest, a night of sleep separated the two sessions (Wagner et al., 2004). Our main objective was to test the hypothesis that task-related cerebral responses induced by implicit learning of the hidden rule already differ during training to the NRT between participants who would eventually gain insight the day after and those who would not. We also aimed at characterizing the changes in brain responses to the NRT that occur between training and retest

which reflect the off-line memory processing leading to insight.

Our results suggest that the functional connectivity of the hippocampus during training materializes the interaction between declarative and procedural motor memory systems and is a crucial parameter that differentiates solver from nonsolver participants.

METHODS

Subjects

Thirty-six healthy young right-handed volunteers (17 women; age: 22 ± 1.76 years) recruited at the University of Liège participated in the experiment. A semistructured interview established the absence of medical, traumatic, psychiatric history, and of sleep disorders. All participants were non-smokers, moderate caffeine and alcohol consumers, and none were on medication or were extreme morning or evening types, as assessed by the Horne and Ostberg Questionnaire (Horne & Ostberg, 1976). None complained of sleep disturbances as determined by the Pittsburgh Sleep Quality Index Questionnaire (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989). Subjects gave written informed consent and received financial compensation for their participation. Their sleep-wake activity was monitored by actigraphy (Cambridge Neuroscience, Cambridge, UK) at least on the 3 days before the experiment and in the night between training and retest. The subjective quality of sleep during the nights preceding the training and retest sessions was checked before the retest session using the St Mary Hospital Sleep Questionnaire (Ellis et al., 1981). The study was approved by the Ethics Committee of the Faculty of Medicine of the University of Liège.

The Number Reduction Task

The same version of the NRT as in previous studies was used (Yordanova, Kolev, Wagner, & Verleger, 2009; Yordanova et al., 2008; Wagner et al., 2004; Rose et al., 2002). In each trial of the task, participants had to process an eight-digit string, generating as a result a string of seven response digits (see Figure 1A). At the beginning of each trial, a different string of eight digits would be presented on the screen. All strings were composed of three different digits, "1," "4," and "9." No other digit could occur. Participants were asked to process the stimuli pairwise. At the beginning of a trial, the first pair of digits consisted of the two automatically generated digits. Thereafter, each pair was composed of the n th computer-generated number and the answer given in response to the $(n - 1)$ th computer-generated number. At each step, the participants had to apply the same-different rule. The "same" rule stated that a combination of two identical digits was associated with the digit itself as an answer. For instance, the combination of the previous answer "1" and the current computer-generated digit "1" implies the response "1" (i.e., "1" and "1" results in "1"; first box on

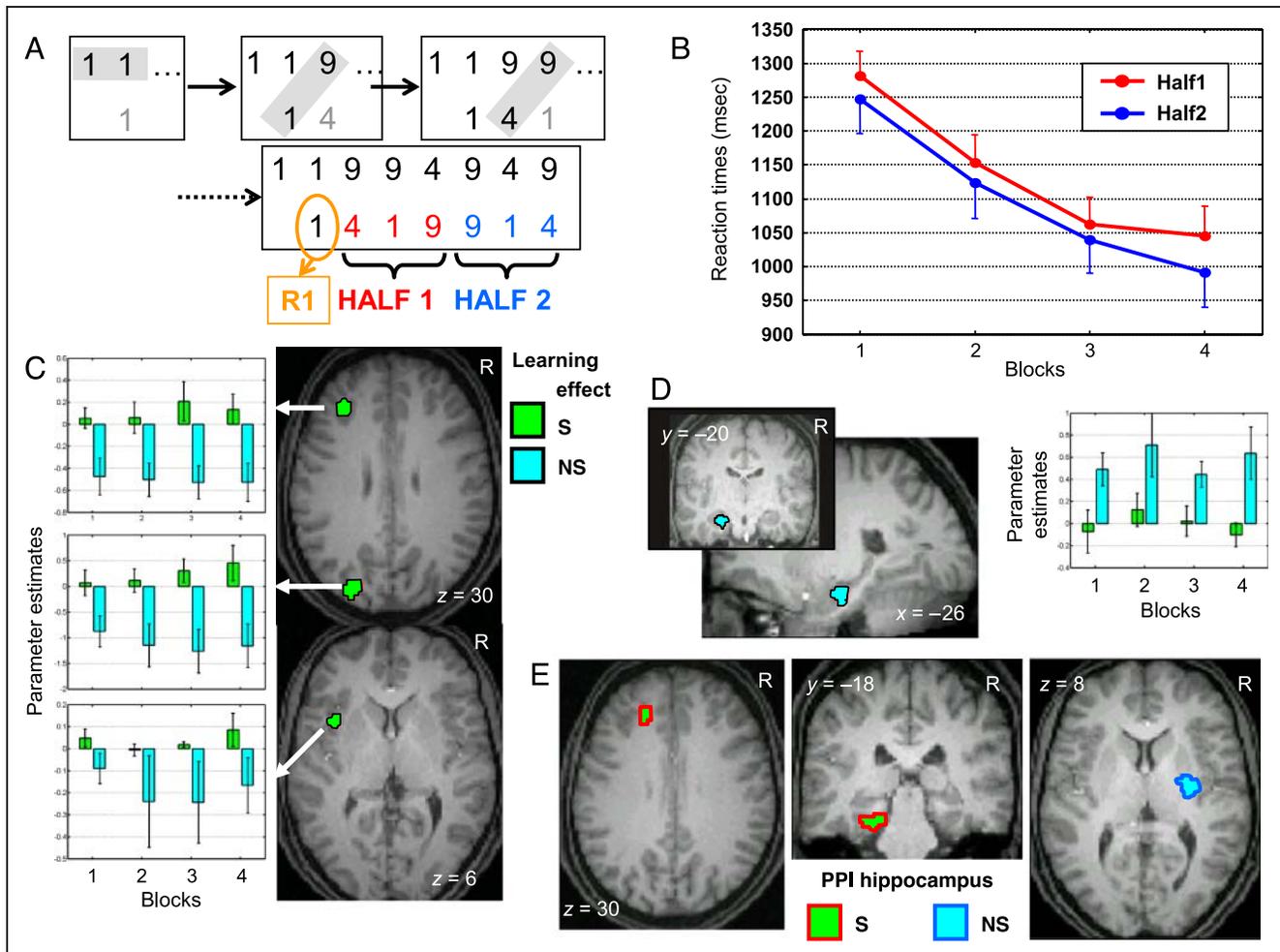


Figure 1. (A) A single trial of the NRT task. After the first pair of computer-generated digits, each subsequent comparison involves the n th computer-generated number and the answer given by the subject in response ($n - 1$)th computer-generated number (shaded numbers) and the participant applies two explicit rules (“same,” “different”; see text) to compute the answers (bottom line). Due to the mirror structure of the last six answers, the last three answers (HALF2) mirror the preceding three answers (HALF1). Importantly, the seventh answer corresponds to the second. A participant who gains insight into this structure will abruptly cut short sequential responding by pressing the solution key immediately after the second response. R1 is the only response that, during retest, is available for analysis both before and after insight. “...” in the upper boxes stands for the digits of the problem which are all displayed at the beginning of the trial but are not represented for the sake of space. (B) Average RTs for HALF1 and HALF2 responses during the 4 blocks of training, across all participants. (C) Larger brain responses in S than in NS, in association with implicit learning (HALF2 > HALF1) during training: superior frontal sulcus, posterior parietal cortex, and insula. Insets show the parameter estimates related to the learning effect (arbitrary units) in S (green bars) and in NS (cyan bars). (D) Larger brain responses in NS than in S, in association with implicit learning (HALF2 > HALF1) during training in the hippocampus. The inset shows the parameter estimates related to the learning effect (arbitrary units) in S (green bars) and in NS (cyan bars). (E) PPI of the hippocampus during training in S (superior frontal sulcus, parahippocampal gyrus, green areas) and NS (putamen, cyan area). Functional statistical results ($p_{\text{uncorrected}} < .001$) are overlaid to the normalized structural image of a typical participant.

Figure 1A). The second rule (“different rule”) states that a pair of two nonidentical digits requires the third digit as an answer. For instance, the combination of the previous answer “1” and a current computer-generated digit “9” leads to the answer “4” (i.e., “1” and “9” results in “4”; second box on Figure 1A). A maximum of 4 sec was offered to respond to any digit pairwise comparison. For each trial, subjects had to determine the digit defined as the “solution digit,” which corresponded to the seventh response. Subjects were instructed that only this solution digit was to be determined and this digit could be entered at any time. The solution digit had to be confirmed by pressing a separate key.

Unknown to the subjects, the digit strings were constructed in such a way that a hidden structure was systematically present in the response pattern: The last three responses (Responses 5–7), that is, the second half of output (HALF2), always mirrored the three previous responses (Responses 2–4), that is, the first half of output (HALF1; Figure 1A). Due to this “hidden structure,” HALF1 responses varied randomly across trials, whereas HALF2 was completely determined by HALF1. This rule also implied that the second response digit of each trial coincided with the final one (i.e., the solution digit). When getting insight into this hidden structure, subjects abruptly cut short

sequential responding by pressing the solution key immediately after the second response. The sudden outbreak of anticipated responses allowed for the estimation of the exact time point at which insight occurred. The instructions regarding the task were presented in the scanner, using a standardized set of sentences in order not to provide any additional information likely to draw the attention to the hidden structure. A delay of 2000 msec always separated the end of a trial from the beginning of the next trial.

Participants were trained to the NRT for 4 blocks in the evening (between 8 and 10 pm) and tested for 7 blocks the next morning (between 8 and 10 am) after a normal night of sleep spent at home. Each practice block comprised 30 trials.

Accuracy was calculated for each subject as mean error rate per session. Mean RT was computed for each of the sequential response (Responses 1 to 7) and block. To parallel the analysis of fMRI data, average RTs for Responses 2 to 4 (HALF1) and Responses 5 to 7 (HALF2) were also computed. The sequence of responses in HALF1 was essentially random, whereas the sequence of responses in HALF2 was imposed by the hidden mirror rule. Only correct responses were considered. Behavioral data were analyzed using a repeated measure ANOVA with response and block as within-subject factors and group (S = solvers; NS = nonsolvers) as between-subject factor.

Postexperimental questionnaires administered after retest first asked whether subjects had the “feeling” that a structure was present in the trials. Subjects were next asked to report the occurrence of relevant thoughts, mentations, or dreams related to the task. “Solvers” were finally asked to indicate the time point at which they gained insight into the hidden rule and if they experienced an abrupt discovery of this rule.

fMRI Data Acquisition and Analysis

Data were acquired during both training and retest, with a 3-T head-only MR scanner (Allegra, Siemens, Erlangen) using a gradient-echo EPI sequence (32 transverse slices, voxel size: $3.4 \times 3.4 \times 3.4$ mm, 30% gap, TR: 2130 msec, TE: 40 msec, FA: 90° , 220 mm FOV). The first three volumes were discarded to account for magnetic saturation effects. A structural MR scan was acquired at the end of the retest (T1-weighted 3-D MP-RAGE sequence, TR: 1960 msec, TE: 4.43 msec, TI: 1100 msec, FOV: 230×173 mm², matrix size $256 \times 192 \times 176$, voxel size: $0.9 \times 0.9 \times 0.9$ mm). Stimuli were displayed on a screen positioned at the rear of the scanner, which the subject could comfortably see through a mirror mounted on the standard head coil.

The fMRI analyses assessed four main functional processes: (1) the implicit learning during training, by contrasting HALF2 to HALF1; (2) the functional connectivity of the hippocampus during training in solvers and nonsolvers, in the context of implicit learning (HALF2–HALF1); (3) the changes in learning-related responses (HALF2–HALF1) from training to testing, indicative of off-line pro-

cessing of implicit memory; (4) the changes in response to R1 occurring in solvers during retest after, relative to before, insight.

fMRI data were analyzed using Statistical Parametric Mapping (SPM2; www.fil.ion.ucl.ac.uk/spm/spm2) implemented in MATLAB 6.1 (Mathworks, Sherborn, MA). Functional scans were realigned using iterative rigid-body transformations that minimize the residual sum of square between the first and subsequent images. They were normalized to the MNI EPI template (2-D spline, voxel size: $2 \times 2 \times 2$ mm) and spatially smoothed with a Gaussian kernel with FWHM of 8 mm.

Functional data were analyzed using a mixed-effects model and conducted in two serial steps, accounting respectively for fixed and random effects. For each subject, a first-level analysis was performed in each voxel using a general linear model. Responses to Inputs 5–7 (the second half of responses, HALF2) that were determined by the hidden rule and responses to Inputs 2–4 (the first half of responses, HALF1) were modeled as box cars with variable duration, convolved with a canonical hemodynamic response function. The duration was computed by the individual RTs for HALF1 and HALF2 for each trial. HALF2 was orthogonalized with respect to HALF1 to account for colinearity between HALF 1 and HALF 2. Movement parameters estimated during realignment (translations in x -, y -, and z -directions and rotations around x -, y -, and z -axes) and a constant vector were also included in the matrix as variables of no interest. High-pass filter was implemented using a cutoff period of 128 sec in order to remove the low-frequency drifts from the time series. Serial autocorrelations were estimated with a restricted maximum likelihood algorithm using an autoregressive model of order 1 (with white noise). Regionally specific condition effects were assessed using linear contrasts testing for differences between HALF2 and HALF1 (i.e., implicit learning of the hidden rule) during training or between training and retest. In the latter case, for retest sessions, only data available before the subjects gained insight into the hidden structure were included in the analysis.

The individual summary statistical images were fed into the second-level analysis which consisted of one-sample t tests checking for the effect of interest separately for each group (either S or NS) and in two-sample t tests comparing the responses between the two groups.

The resulting set of voxel values for each contrast constituted maps of the t statistics [SPM(T)], thresholded at $p < .001$. Statistical inferences were performed at $p_{\text{svc}} < .05$, after correction for multiple comparisons on small spherical volumes centered on published coordinates of interest (svc; 10 mm radius; see Table 2). When indicated, exclusive masks were used, thresholded at $p < .05$, uncorrected. This procedure identifies the responses that remain significant in one group (e.g., S) after exclusion of the responses found in the other (e.g., NS). In other words, exclusive masking identifies the responses selectively significant for the first group.

A similar procedure was followed to estimate the change in learning-related response (HALF2–HALF1) between the last block of training and the first block of retest as well as the changes in responses to R1 (the first of the 7 responses) that occur at retest in solvers, during the block in which participants gained insight into the hidden rule.

Psychophysiological interaction (PPI) analyses were computed to test the hypothesis that functional connectivity between the hippocampus and the rest of the brain not only differed between HALF1 and HALF2 but also between solvers and nonsolvers during training. For each individual, the coordinates of the seed areas corresponded to the local maxima detected within 10 mm of the peak voxel of the random effects analysis reported above (coordinates found in NS). A new linear model was prepared for PPI analyses at the individual level, using three regressors. One regressor represented the HALF1 and HALF2 blocks. The second regressor was the activity in the hippocampus. The third regressor represented the interaction of interest between the first (psychological) and the second (physiological) regressor. To build this third regressor, the underlying neuronal activity was first estimated by a parametric empirical Bayes formulation, combined with the psychological factor and, subsequently, convolved with the hemodynamic response function (Gitelman, Penny, Ashburner, & Friston, 2003). The model also included movement parameters. Individual summary statistic images obtained at the first-level (fixed effects) analysis were spatially smoothed (6 mm FWHM Gaussian kernel) and entered in a second-level (random effects) analysis using two-sample *t* tests to compare the functional connectivity between groups. Inferences were conducted as for the main effect analysis.

RESULTS

Demographic Data

Four subjects gained insight during initial training and were not retested. Sixteen participants gained insight at retest (S). In most cases, insight occurred during the second block (median = 1.5, range 1 to 7). Eight of the solvers gained insight in the first block of the retest, suggesting that it was not induced by an extensive repeated exposure to the task but to a predisposition available before retest. Postexperimental questionnaires confirmed the gain of explicit knowledge of the hidden rule occurred in all solvers during retest and not during training. Sixteen subjects never discovered the hidden structure (NS). None of them had the “feeling that a structure was present in the trials.” No subject reported to have had thoughts, mentations, or dreams related to the task between training and retest.

Sleep Data

Subjective sleep durations as indicated by the St Mary Hospital Sleep Questionnaire were not significantly different between solver and nonsolver subjects on the night preced-

ing the training session (mean \pm SEM: S = 9 ± 0.9 hr, NS = 8.6 ± 2.2 hr) and the night between training and retest sessions (S = 8.7 ± 0.6 hr, NS = 9.2 ± 0.8 hr).

No significant main effects of group [$F(1, 30) = 0.07$, $p = .78$] or night [$F(1, 30) = 0.23$, $p = .64$] or Group \times Night interaction [$F(1, 30) = 1.74$, $p = .19$] were found for mean sleep durations. ANOVA on actigraphic data on the nights preceding training and the night between training and retest showed a trend for the effect of night [$F(1, 30) = 3.07$, $p = .09$] but did not show any significant effect of group [$F(1, 30) = 0.01$, $p = .91$] or any significant Group \times Night interaction [$F(1, 30) = 1.51$, $p = .23$].

Behavioral Data

During training, no difference in error rates was observed between solvers and nonsolvers [$F(1, 30) = 5.3$, $p = .87$]. Error rates generally decreased across blocks [$F(3, 90) = 32.8$, $p < .001$], mainly between the first and second blocks (mean error rate = 21.6%, 5.4%, 4.2%, 3.8% for Blocks 1 to 4).

RTs likewise decreased across blocks [$F(3, 90) = 76.5$, $p < .001$] and differed between response position (Responses 1 to 7) [$F(6, 180) = 65.3$, $p < .001$] (Figure 1B). RTs decreased differently for the various response positions as training progressed, indicated by a significant interaction between block and response position [$F(18, 540) = 2.57$, $p < .001$]. As described below, this interaction pattern mainly resulted from a decrease in RTs for the last responses of each problem (Responses 6 and 7). Solver and nonsolver subjects did not differ in terms of RTs during training: There was no group effect [$F(1, 30) = 0.39$, $p = .54$] or any significant Block \times Group [$F(3, 90) = 0.37$, $p = .77$] or Block \times Response \times Group interaction [$F(18, 540) = 0.87$, $p = .62$]. To parallel the fMRI analysis, we also contrasted the last three responses, determined by the task structure (Positions 5–7, HALF2) to the three preceding responses (Positions 2–4, HALF1), calculating for both response types mean RTs of the three respective responses (Rose et al., 2002). Indicating implicit learning of the hidden rule (Rose et al., 2002), RTs were generally shorter for HALF2, as compared to HALF1 across blocks [$F(1, 30) = 21.9$, $p < .001$] (Figure 1B), but not differently so between solvers and nonsolvers [$F(1, 30) = 0.06$, $p = .80$, for interaction]. To assess the effect of the salient feature associated with direct repetition of responses in Positions 4 and 5, we compared Responses 2 and 3 to Responses 6 and 7. RTs tended to decrease more rapidly for Responses 6–7 than 2–3 from Blocks 1 to 4 [$F(1, 30) = 2.74$, $p = .09$]. In contrast, there was no significant difference in the change in RT between Responses 4 and 5 throughout training [$F(1, 30) = 0.64$, $p = .43$]. These findings indicate that in the course of training, the development of sensitivity to the mirror structure is not simply driven by the response positions that are immediately repeated. No significant difference was observed between solvers and nonsolvers, in

terms of RTs for Positions 2 and 3, relative to Positions 6 and 7 [$F(1, 30) = 0.34, p = .57$].

In a second analysis, we identified changes in RTs between the last training block and the first retest block. RTs were globally shorter during retest than training [$F(1, 30) = 15.26, p < .001$], but not significantly so between solvers and nonsolvers [$F(1, 30) = 0.001, p = .97$]. There was no significant change in RT for HALF2 relative to HALF1 between training and retest [$F(1, 30) = 0.99, p = .33$]. However, there was a significant difference in the change in RT for HALF2 relative to HALF1 between solvers and nonsolvers [$F(1, 30) = 5.05, p = .03$]. The decrease in RT for HALF2 relative to HALF1 between training and retest was significant in nonsolvers [$F(1, 30) = 5.25, p = .03$] but not in solvers [$F(1, 30) = 0.78, p = .38$].

A final analysis showed that there was no significant difference in RTs to R1 after relative to after insight during retest (before: 1215 ± 270 msec; after: 1282 ± 345 msec, $p > .05$).

fMRI Data

First, the neural correlates of implicit learning of the hidden structure were assessed by contrasting the responses to HALF2 and HALF1 during training. Across all participants, responses associated with implicit learning (HALF2 > HALF1) were observed in a large number of bilateral areas: medial and lateral prefrontal areas, anterior and posterior cingulate cortex, inferior parietal cortex, middle temporal gyrus, posterior insula, cuneus, and parahippocampal gyrus (Table 1).

There were significant differences in learning-related responses between solvers and nonsolvers (Table 2). Learning-related responses were significantly larger in solvers than in nonsolvers, in the left superior frontal sulcus (SFS), left posterior parietal cortex (angular gyrus), and in the left anterior insula (Figure 1C). In contrast, a larger learning effect was detected in nonsolvers than in solvers, in the left anterior hippocampus (Figure 1D). Moreover, the analysis of functional connectivity revealed that the left hippocampus established distinctive functional connections with other brain regions, in the context of learning (HALF2 relative to HALF1), depending on the participant's final status as solver or nonsolver (Figure 1E). The hippocampus was functionally connected with the right putamen and the left putamen/globus pallidus, in nonsolvers but not in solvers (exclusive masking by S). In contrast, in solvers, but not in nonsolvers (exclusive masking by NS), the hippocampus was functionally related to the left parahippocampal gyrus and the left SFS. The latter region was located next to the frontal area detected in the differential learning effect previously reported between solvers and nonsolvers.

Second, we examined how memory traces were processed overnight. We characterized the changes in learning-related responses between the last block of training and the first block of retest, which are deemed to reflect best the off-line processes taking place during the posttrain-

Table 1. Responses Associated with Implicit Rule Learning (HALF2 > HALF1) during Training across All Participants

Area	L/R	x	y	z	Z Score	p_{cor}
Lateral orbito-frontal cortex	L	-36	18	-20	7.33	<.001
	R	30	20	-22	7.55	<.001
Ventrolateral prefrontal cortex	L	-50	34	-12	6.87	<.001
	R	48	40	-8	6.63	<.001
Dorsolateral prefrontal cortex	L	-44	18	52	5.53	.003
	R	46	20	52	5.74	<.001
Medial prefrontal cortex	L	-8	48	8	7.73	<.001
	R	6	44	22	7.67	<.001
Inferior parietal lobule	L	-58	-54	44	>10	<.001
	R	58	-52	38	7.94	<.001
Middle temporal gyrus	L	-64	-46	-4	6.47	<.001
	R	66	-42	-4	6.15	<.001
Cuneus	L	-14	-100	26	7.44	<.001
	R	12	-92	34	5.87	<.001
Posterior Insula	L	-38	-8	-12	5.20	.004
	R	38	-10	-4	6.14	<.001
Anterior cingulate cortex	L	-8	40	22	>10	<.001
	R	4	40	14	>10	<.001
Posterior cingulate cortex	L	-2	-24	38	>10	<.001
	R	8	-38	34	7.72	<.001
Parahippocampal gyrus	L	-28	-36	-20	7.84	<.001
	R	22	40	-8	7.16	<.001

A single peak voxel is reported by brain area. L/R = left or right; x, y, z = coordinates (mm) in the stereotactic space defined by the Montreal Neurological Institute. p_{cor} = probability of response after correction for multiple comparisons over the whole brain.

ing period (Figure 2A, pink; Table 2). During the first retest block, we considered only the trials in which there is no objective evidence that participants gained insight into the hidden rule. The comparison between late training and early retest is thus warranted. Larger responses at retest were observed in ventral medial prefrontal cortex (vmPFC) and the left putamen, in solvers but not in nonsolvers (exclusive mask). Finally, for each solver, we characterized the neural responses to the first response of each trial at retest before and after insight (R1, Figure 1A). Responses were larger before insight, when the hidden structure was still preconscious (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006), than after insight in anterior vmPFC, supplementary motor area, and left primary motor cortex (Figure 2A, blue; Table 3). vmPFC overlapped with the area in which responses significantly increased overnight (Figure 2A, violet). When subjects gain insight, without any change in stimulus or instructions, a distributed

network recruited the anterior and posterior cingulate cortices, precuneus, and bilateral frontal areas and parietal areas (Figure 2B; Table 3).

DISCUSSION

In this study, we used fMRI to determine neural precursors of sleep-associated insight in the NRT. Normal volunteers practiced the NRT during two fMRI sessions (training and retest), separated by a night of sleep. The results confirm our hypothesis that the memories generated during training already differ at the systems level between participants who will (S) or those who will not (NS) subsequently gain insight into the hidden rule during retest. Implicitly learning the hidden rule recruits the hippocampus in nonsolvers and a set of cortical areas in solvers. In nonsolvers, the hippocampus functionally interacts with the basal ganglia during training. In contrast, in solvers, the activity of the hippocampus increases with implicit learning only in proportion to frontal responses. Furthermore, in solvers, the changes in learning-related responses overnight are detected in

vmPFC, an area previously implicated in the consolidation of declarative memory. During retest, before insight, this region continues to respond in association to learning. After insight, a distributed set of parietal, frontal, and temporal areas supports the optimal behavior.

Neural Correlates of Implicit Learning Differ between Solvers and Nonsolvers Already at Training

The main finding of this study consists of the difference in the neural correlates of implicit learning detected during training between solvers and nonsolvers, in the absence of significant behavioral differences. Similarly, differences in EEG-evoked responses were reported between solvers and nonsolvers during initial training to the NRT (Lang et al., 2006). However, these differences appeared as individual EEG trait characteristics because they were detected at the very first block of practice, before any learning, whereas the difference in fMRI responses reported here are related to implicit learning (differences between HALF2 and HALF1).

Table 2. Group-selective fMRI Responses

Area	L/R	x	y	z	Z Score	p_{SVC}
<i>Learning Effect—Training—S > NS</i>						
Superior frontal sulcus (Chein & Schneider, 2005)	L	-32	34	30	3.14	.030
Anterior insula (Chein & Schneider, 2005)	L	-38	10	6	3.33	.018
Posterior parietal cortex (Chein & Schneider, 2005)	L	-24	-84	28	3.11	.032
<i>Learning Effect—Training—NS > S</i>						
Hippocampus (Schendan et al., 2003)	L	-26	-20	-22	3.47	.012
<i>Functional Connectivity of the Hippocampus—Training—NS (Exclusive Mask: S)</i>						
Putamen (Penhune & Doyon, 2002)	R	30	-10	4	4.13	.002
Putamen/Globus pallidus (Penhune & Doyon, 2002)	L	-26	-14	8	3.12	.043
<i>Functional Connectivity of the Hippocampus—Training—S (Exclusive Mask: NS)</i>						
Superior frontal sulcus (Sakai et al., 1998)	L	-22	38	32	3.31	.046
Parahippocampal gyrus (Schendan et al., 2003)	L	-24	-18	-24	3.24	.036
<i>Increases in Response from Training to Retest—S (Exclusive Mask: NS)</i>						
Ventral medial prefrontal cortex (Takashima et al., 2006)	R	10	32	-16	3.20	.037
Putamen (Penhune & Doyon, 2002)	L	-32	4	8	3.73	.023
<i>Changes in Response from Training to Retest—NS</i>						
No significant results						

L/R = left or right; x, y, z = coordinates (mm) in the stereotactic space defined by the Montreal Neurological Institute; p_{SVC} = probability of response after correction for multiple comparisons over small volumes of interest defined in the literature (reference in Column 1).

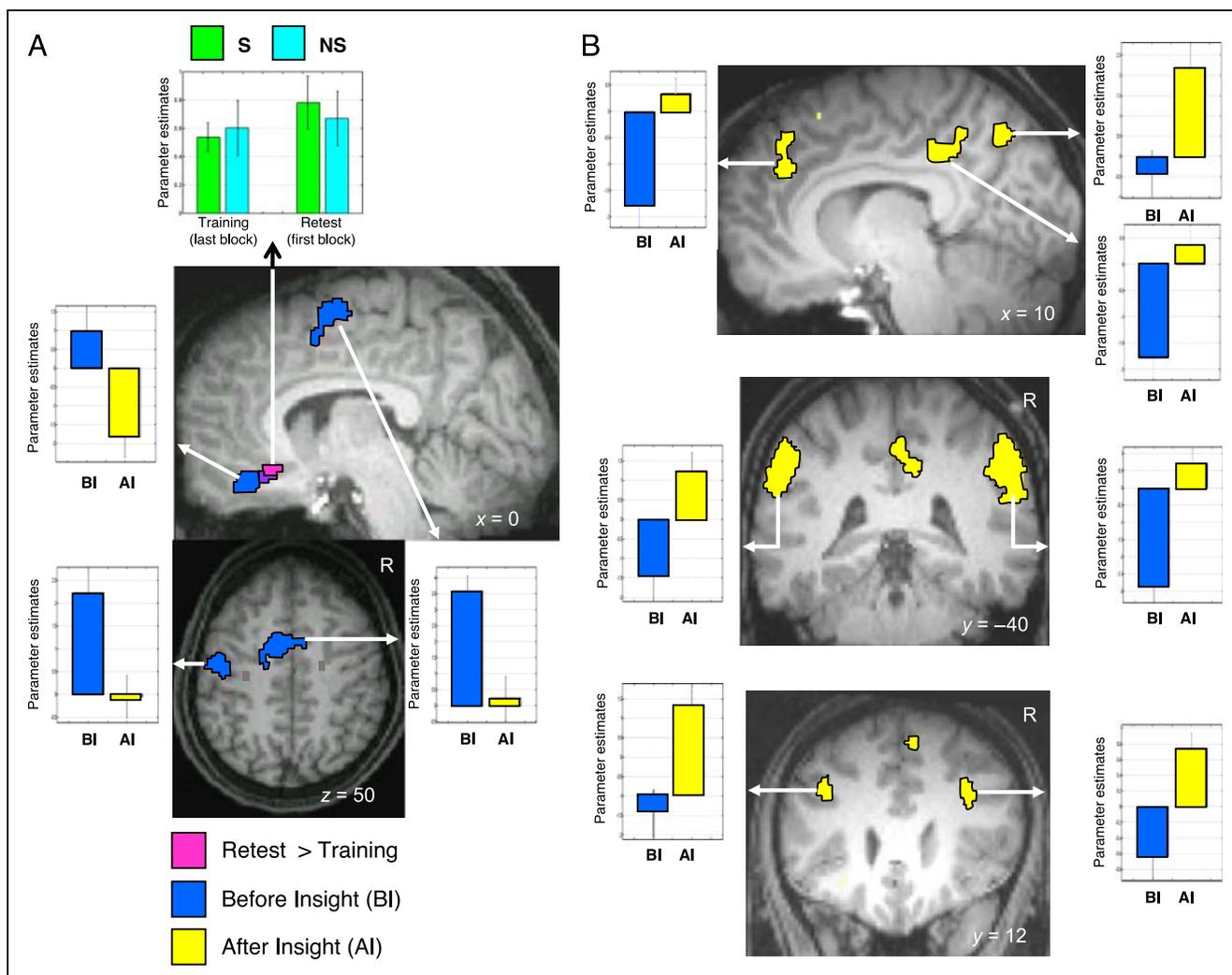


Figure 2. (A) This panel combines the results of two analyses for solvers. (1) Enhanced overnight response in vmPFC (pink). (2) Brain responses to R1 larger before than after insight in vmPFC, supplementary area, and left primary motor cortex (blue). In violet, responses detected in analyses (1) and (2) overlap in vmPFC. (B) During retest, the responses to R1 are larger after than before insight in a distributed set of frontal and parietal areas. Functional statistical results ($p_{\text{uncorrected}} < .001$) are overlaid to the normalized structural image of a typical participant.

In solvers, implicit learning was associated with significant responses in dorsolateral prefrontal cortex, posterior parietal cortex, and insula. These areas belong to a domain-general system mediating controlled processes which support early learning and novice performance (Chein & Schneider, 2005). The results suggest that solvers resorted to a larger extent to controlled processes than nonsolvers to learn the task, potentially allowing a closer monitoring of the task structure. In contrast, larger learning effects were detected in nonsolvers than in solvers in the left anterior hippocampus. Although consistent with the implication of the medial temporal lobe in implicit motor sequence learning (hippocampus: Albouy et al., 2008; Schendan et al., 2003; perirhinal cortex: Rose et al., 2002), this finding was intriguing given that the hippocampus is classically associated with explicit learning in the amnesia literature (Spiers, Maguire, & Burgess, 2001; Squire & Zola, 1996) and, consequently, would intuitively be thought to promote insight (Luo & Niki, 2003).

This finding led us to assess the functional connectivity of the hippocampus during training in the two groups. In solvers, the left hippocampus established a distinctive functional connectivity with dorsolateral prefrontal cortex, in an area close to the region showing a differential learning effect between solvers and nonsolvers. This finding suggests that in solvers, the hippocampus was influenced by its interactions with cortical controlled processes supporting learning. In contrast, in nonsolvers, the hippocampus was functionally connected with the bilateral putamen. The increased putaminal response is consistent with the involvement of cortico-striatal loops in procedural motor skill learning (Albouy et al., 2008; Doyon, Penhune, & Ungerleider, 2003; Peigneux et al., 2000). However, the functional relationships between the medial temporal lobe and the striatum during implicit learning are usually deemed competitive. Probabilistic classification learning is associated with a decreased response in the medial temporal lobe structures, contrasting with an activation of the

basal ganglia (Seger & Cincotta, 2006; Poldrack et al., 2001; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999). During oculomotor sequence learning, a negative functional coupling is observed between the hippocampus and the putamen (Albouy et al., 2008). This competing interaction turns into cooperation between the two memory systems only after a night of sleep (Albouy et al., 2008). Here, we see that implicit learning of the hidden rule is associated with an immediate cooperative interaction between these memory systems in nonsolvers. It is tempting to speculate that this cooperation, associated with fairly good implicit performance levels, leads nonsolvers to stick to the implicit strategy rather than to try alternative solutions.

Off-line Memory Processing Underlying Subsequent Insight

Another important finding consists of the changes in responses associated with implicit learning from training to retest. In the absence of practice between sessions, these changes reflect memory processes that are taking place

off-line (here by design, during a night of sleep). Whereas no such modifications were observed in nonsolvers, responses in vmPFC significantly increased overnight in solvers. Intriguingly, vmPFC is classically involved in systems consolidation of declarative, rather than procedural, memories. Although it is usually associated with long-term consolidation of declarative memories (Gais et al., 2007; Takashima et al., 2006), it was shown to be recruited as early as 72 hr after encoding for emotional memories (Sterpenich et al., 2007). The present data show that, in the case of the NRT, the off-line processing of implicit learning of the hidden rule is associated with enhanced vmPFC responses in participants who will eventually gain insight during retest, only 12 hr after training.

Intriguingly, the responses in the putamen were significantly increased overnight in solvers but not in nonsolvers. This result contrasts with earlier reports showing enhanced striatal responses overnight after *implicit* sequence learning (Albouy et al., 2008). However, consistent with other reports (Sterpenich et al., 2007; Orban et al., 2006), this result suggests that off-line memory processing, including

Table 3. Changes in Responses to R1 after Relative to before Insight during Retest in S

Area	L/R	<i>x</i>	<i>y</i>	<i>z</i>	<i>Z Score</i>	<i>p_{SVC}</i>
<i>Response to R1: Before > After Insight</i>						
Medial prefrontal cortex (Takashima et al., 2006)		0	32	-18	4.2	.001
Supplementary motor area (Doyon et al., 2002)	L	-4	-2	60	3.96	.007
Primary motor cortex (Penhune & Doyon, 2002)	L	-48	-6	48	3.27	.047
<i>Response to R1: After > Before Insight</i>						
Anterior cingulate cortex (Henson, Rugg, Shallice, Josephs, & Dolan, 1999)	L	-6	10	28	3.95	.007
Dorsal medial prefrontal cortex (Henson et al., 1999)	R	10	40	34	3.64	.017
Dorsolateral prefrontal cortex (Marois, Yi, & Chun, 2004; Willingham, Salidis, & Gabrieli, 2002)	R	36	28	36	5.07	
	R	48	10	40	3.65	.017
	R	50	13	35	3.52	.024
	R	42	12	42	4.49	
	L	-14	12	54	5.05	
Posterior cingulate cortex (Willingham et al., 2002; Henson et al., 1999)	R	8	-34	30	3.46	.029
	L	-6	-32	38	3.19	.050
Precuneus (Henson et al., 1999)	R	6	-66	52	3.73	.014
Inferior parietal lobule (Henson et al., 1999)	L	-54	-42	44	3.51	.025
	R	60	-44	40	5.29	.000
Intraparietal sulcus (Marois et al., 2004)	R	34	-54	42	3.37	.037
Temporo-parietal junction (Marois et al., 2004)	R	58	-40	26	3.55	.022
	L	-58	-34	30	3.27	.046

L/R = left or right; *x*, *y*, *z* = coordinates (mm) in the stereotactic space defined by the Montreal Neurological Institute; *p_{SVC}* = probability of response after correction for multiple comparisons over small volumes of interest defined in the literature (reference in Column 1).

those influenced by sleep, can occur simultaneously in distinct neural circuits.

The Role of Sleep on Insightful Behavior : A Note of Caution

Insight of the NRT hidden rule occurs significantly more frequently across early-night sleep, rich in deep non-rapid eye movement (NREM) sleep, than across late-night sleep, provided that a certain level of implicit knowledge is acquired before sleep (Yordanova et al., 2008). In the framework of the systems consolidation of declarative memories, the recruitment of vmPFC is usually believed to arise from its functional interactions with the hippocampus, especially during NREM sleep (Buzsaki, 1996). For instance, a sleep-dependent functional interaction between the hippocampus and vmPFC was observed in humans during a recall session taking place 48 hr following encoding (Gais et al., 2007). Collectively, these data would suggest that hippocampus-dependent, still implicit, representations of the hidden rule are processed during NREM sleep, through interactions between the hippocampus and vmPFC. However, in the present study, in the absence of an awake control group, we cannot specifically assess the effects of sleep of the emergence of insight. In addition, we did not observe any change in hippocampal responses from training to retest, and we have no positive evidence that the vmPFC responses were enhanced overnight by interactions with the hippocampus.

Emergence of Insight during Retest

During retest, responses associated with the first response (R1) were observed in vmPFC, the supplementary motor area, and left primary motor cortex. The latter two areas are obviously involved in execution of learned motor sequences (Orban et al., 2010). The recruitment of vmPFC can be interpreted in two different ways. The data discussed above would link this response to both the application of the implicitly learned rule and off-line memory processing. Alternatively, the response recorded in vmPFC might be related to reward signals associated with the accurate execution of the NRT following implicit learning (Balleine, Delgado, & Hikosaka, 2007).

After insight, responses to R1 profoundly change and recruit a set of widespread associative frontal, cingulate, and parietal regions. In particular, the precuneus, due to its extensive anatomical connectivity that links major structural modules of brain connectivity (Hagmann et al., 2008), has been associated with conscious processes, some of which are relevant to gaining insight into the hidden structure, such as memory retrieval or mental imagery in deductive reasoning and sequential finger movements (Cavanna & Trimble, 2006). The recruitment of this distributed cortical network would instantiate a distinct internal space where information can be shared across a broad variety of processes (Dehaene et al., 2006). The ignition of this

network would be associated to the emergence of insight (Dehaene & Changeux, 2005).

Conclusions

The present data indicate that a difference in the neural correlates of implicit learning exists between solvers and nonsolvers already during the initial exposure to the NRT, suggesting that the early structuring of the memory trace determines its subsequent fate during consolidation and, ultimately, conditions the emergence of insight. In this process, the hippocampus seems to play a pivotal role in biasing participants' strategy toward implicit or explicit memory processes, depending on its functional connectivity with either basal ganglia or associative prefrontal regions. A plausible scenario suggests that the overnight enhancement of responses in vmPFC observed in solvers results from its interplay with the hippocampus during sleep. The next morning, insight is associated with the recruitment of a widespread set of polymodal associative areas within which implicit knowledge of the hidden rule can be shared, thereby reaching consciousness.

Acknowledgments

This research was supported by the Belgian National Fund for Scientific Research (FNRS Belgique), Queen Elisabeth Medical Foundation, Research Fund of the University of Liège, and "Inter-university Attraction Poles Programme–Belgian State–Belgian Science Policy." A. D., E. B., M. D., V. S., G. V., T. D. V., F. C., M. B., and P. M. are supported by the FNRS. A. D. was supported by the Fondation Fyssen.

Reprint requests should be sent to Pierre Maquet, Cyclotron Research Centre (B30), 6, Allée du 8 Août, University of Liège, Belgium, or via e-mail: pmaquet@ulg.ac.be, Web: www.ulg.ac.be/crc.

REFERENCES

- Albouy, G., Sterpenich, V., Balteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., et al. (2008). Both the hippocampus and striatum are involved in consolidation of motor sequence memory. *Neuron*, *58*, 261–272.
- Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience*, *27*, 8161–8165.
- Buyse, D. J., Reynolds, C. F., III, Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*, 193–213.
- Buzsaki, G. (1996). The hippocampo-neocortical dialogue. *Cerebral Cortex*, *6*, 81–92.
- Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*, 564–583.
- Chein, J. M., & Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Brain Research, Cognitive Brain Research*, *25*, 607–623.

- Dehaene, S., & Changeux, J. P. (2005). Ongoing spontaneous activity controls access to consciousness: A neuronal model for inattentional blindness. *PLoS Biology*, *3*, e141.
- Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*, 204–211.
- Doyon, J., & Benali, H. (2005). Reorganization and plasticity in the adult brain during learning of motor skills. *Current Opinion in Neurobiology*, *15*, 161–167.
- Doyon, J., Penhune, V., & Ungerleider, L. G. (2003). Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*, *41*, 252–262.
- Doyon, J., Song, A. W., Kami, A., Lalonde, F., Adams, M. M., & Ungerleider, L. G. (2002). Experience-dependent changes in cerebellar contributions to motor sequence learning. *Proceedings of the National Academy of Sciences, U.S.A.*, *99*, 1017–1022.
- Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*, 109–120.
- Ellis, B. W., Johns, M. W., Lancaster, R., Raptopoulos, P., Angelopoulos, N., & Priest, R. G. (1981). The St. Mary's Hospital sleep questionnaire: A study of reliability. *Sleep*, *4*, 93–97.
- Gais, S., Albouy, G., Boly, M., Dang-Vu, T. T., Darsaud, A., Desseilles, M., et al. (2007). Sleep transforms the cerebral trace of declarative memories. *Proceedings of the National Academy of Sciences, U.S.A.*, *104*, 18778–18783.
- Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. *Neuroimage*, *19*, 200–207.
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, *6*, e159.
- Haider, H., & Rose, M. (2007). How to investigate insight: A proposal. *Methods*, *42*, 49–57.
- Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, *19*, 3962–3972.
- Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms. *International Journal of Chronobiology*, *4*, 97–110.
- Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biology*, *2*, E97.
- Lang, S., Kanngieser, N., Jaskowski, P., Haider, H., Rose, M., & Verleger, R. (2006). Precursors of insight in event-related brain potentials. *Journal of Cognitive Neuroscience*, *18*, 2152–2166.
- Luo, J., & Niki, K. (2003). Function of hippocampus in “insight” of problem solving. *Hippocampus*, *13*, 316–323.
- Maquet, P., & Ruby, P. (2004). Psychology: Insight and the sleep committee. *Nature*, *427*, 304–305.
- Marois, R., Yi, D. J., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron*, *41*, 465–472.
- Orban, P., Peigneux, P., Lungu, O., Albouy, G., Breton, E., Laberenne, F., et al. (2010). The multifaceted nature of the relationship between performance and brain activity in motor sequence learning. *Neuroimage*, *49*, 694–702.
- Orban, P., Rauchs, G., Baletau, E., Degueldre, C., Luxen, A., Maquet, P., et al. (2006). Sleep after spatial learning promotes covert reorganization of brain activity. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 7124–7129.
- Peigneux, P., Maquet, P., Meulemans, T., Destrebecqz, A., Laureys, S., Degueldre, C., et al. (2000). Striatum forever, despite sequence learning variability: A random effect analysis of PET data. *Human Brain Mapping*, *10*, 179–194.
- Penhune, V. B., & Doyon, J. (2002). Dynamic cortical and subcortical networks in learning and delayed recall of timed motor sequences. *Journal of Neuroscience*, *22*, 1397–1406.
- Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature*, *414*, 546–550.
- Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*, 245–251.
- Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology*, *13*, 564–574.
- Poldrack, R. A., & Rodriguez, P. (2004). How do memory systems interact? Evidence from human classification learning. *Neurobiology of Learning and Memory*, *82*, 324–332.
- Rose, M., Haider, H., Weiller, C., & Buchel, C. (2002). The role of medial temporal lobe structures in implicit learning: An event-related fMRI study. *Neuron*, *36*, 1221–1231.
- Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Sasaki, Y., & Putz, B. (1998). Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. *Journal of Neuroscience*, *18*, 1827–1840.
- Schacter, D. L., & Tulving, E. (Eds.) (1994). *Memory systems 1994*. Cambridge, MA: MIT Press.
- Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, *37*, 1013–1025.
- Seger, C. A., & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral Cortex*, *16*, 1546–1555.
- Spiers, H. J., Maguire, E. A., & Burgess, N. (2001). Hippocampal amnesia. *Neurocase*, *7*, 357–382.
- Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences, U.S.A.*, *93*, 13515–13522.
- Sterpenich, V., Albouy, G., Boly, M., Vandewalle, G., Darsaud, A., Baletau, E., et al. (2007). Sleep-related hippocampo-cortical interplay during emotional memory recollection. *PLoS Biology*, *5*, e282.
- Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., et al. (2006). From the cover: Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 756–761.
- Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, *427*, 352–355.
- Willingham, D. B., Salidis, J., & Gabrieli, J. D. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *Journal of Neurophysiology*, *88*, 1451–1460.
- Yordanova, J., Kolev, V., Verleger, R., Bataghva, Z., Born, J., & Wagner, U. (2008). Shifting from implicit to explicit knowledge: Different roles of early- and late-night sleep. *Learning and Memory*, *15*, 508–515.
- Yordanova, J., Kolev, V., Wagner, U., & Verleger, R. (2009). Covert reorganization of implicit task representations by slow wave sleep. *PLoS One*, *4*, e5675.