Report

Current Biology

Direct Electrophysiological Evidence for Prefrontal Control of Hippocampal Processing during Voluntary Forgetting

Highlights

- We investigated iEEG oscillations in DLPFC and hippocampus of epilepsy patients
- Voluntary forgetting increases DLPFC 3–5 Hz power and hippocampal 6–18 Hz power
- Three complementary directed connectivity metrics reveal DLPFC→HC coupling
- Directional DLPFC→HC information flow is specific for to-be-forgotten items

Authors

Carina R. Oehrn, Juergen Fell, Conrad Baumann, ..., Henrik Kessler, Simon Hanslmayr, Nikolai Axmacher

Correspondence

nikolai.axmacher@rub.de

In Brief

Using direct electrophysiological recordings, Oehrn et al. show that voluntary forgetting depends on enhanced theta oscillations in DLPFC and low-frequency activity in hippocampus. Converging evidence from time-lagged power correlations, phase synchronization, and Granger causality analysis indicates a top-down control of hippocampus by the DLPFC.

Oehrn et al., 2018, Current Biology *28*, 1–7
September 24, 2018
https://doi.org/10.1016/j.cub.2018.07.042

CellPress

Please cite this article in press as: Oehrn et al., Direct Electrophysiological Evidence for Prefrontal Control of Hippocampal Processing during Voluntary Forgetting, Current Biology (2018), https://doi.org/10.1016/j.cub.2018.07.042

Current Biology
Report

Direct Electrophysiological Evidence for Prefrontal Control of Hippocampal Processing during Voluntary Forgetting

Carina R. Oehrn,[1] Juergen Fell,[2] Conrad Baumann,[3] Timm Rosburg,[4] Eva Ludowig,[5] Henrik Kessler,[6] Simon Hanslmayr,[7,8] and Nikolai Axmacher[9,10,*]
[1]Department of Neurology, University Hospital of Giessen and Marburg, Baldingerstrasse, 35043 Marburg, Germany
[2]Department of Epileptology, University of Bonn, Sigmund-Freud-Strasse 25, 53127 Bonn, Germany
[3]Median Medical Center Mecklenburg, Blumenstrasse 3, 19217 Vitense, Germany
[4]Department of Forensic Psychiatry, University Psychiatric Clinics Basel, Wilhelm Klein-Strasse 27, 4002 Basel, Switzerland
[5]Knowledge2Brain, Kaaplandstraat 69, 6543PC Nijmegen, the Netherlands
[6]Department of Psychosomatic Medicine and Psychotherapy, LWL-University Hospital, Ruhr University Bochum, Alexandrinenstrasse 1-3, 44791 Bochum, Germany
[7]School of Psychology, University of Birmingham, Edgbaston, B15 2TT, UK
[8]Centre for Human Brain Health, University of Birmingham, Edgbaston, B15 2TT, UK
[9]Department of Neuropsychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Universitaetsstrasse 150, 44801 Bochum, Germany
[10]Lead Contact
*Correspondence: nikolai.axmacher@rub.de
https://doi.org/10.1016/j.cub.2018.07.042

SUMMARY

Forgetting does not necessarily reflect failure to encode information but can, to some extent, also be voluntarily controlled. Previous studies have suggested that voluntary forgetting relies on active inhibition of encoding processes in the hippocampus by the dorsolateral prefrontal cortex (DLPFC) [1–4]. During attentional and sensorimotor processing, enhanced DLPFC theta power alongside increased alpha/beta oscillations are a neural signature of an inhibitory top-down mechanism, with theta oscillations reflecting prefrontal control and alpha/beta oscillations occurring in areas targeted by inhibition [5–12]. Here, we used intracranial EEG recordings in presurgical epilepsy patients implanted in DLPFC (n = 13) and hippocampus (n = 15) during an item-method directed forgetting paradigm. We found that voluntary forgetting is associated with increased neural oscillations in the low theta band (3–5 Hz) in DLPFC and in a broad theta/alpha/beta (6–18 Hz) frequency range in hippocampus. Combining time-lagged correlation analysis, phase synchronization, and Granger causality in 6 patients with electrodes in both DLPFC and hippocampus, we obtained converging evidence for a top-down control of hippocampal activity by the DLPFC. Together, our results provide strong support for a model in which voluntary forgetting relies on enhanced inhibition of the hippocampus by the DLPFC.

RESULTS AND DISCUSSION

Hippocampal and Prefrontal Oscillations

Memory for to-be-forgotten (TBF) words was worse than for to-be-remembered (TBR) words (Figures 1 and S1A). Electrophysiological data were analyzed by means of non-parametric cluster-based permutation statistics free of any *a priori* assumptions regarding specific frequency bands or time periods (STAR Methods [13]). To this end, we entered all time (0–1.5 s after cue onset) and frequency points (2–29 Hz) into the analysis.

When compared to TBR cues, TBF cues elicited significantly enhanced iEEG activity in the low-theta range (3–5 Hz) in the dorsolateral prefrontal cortex (DLPFC) (Figures 1B and 2A; 568–1,058 ms; $p = 0.041$), and in a broader theta/alpha/beta frequency range (6–18 Hz) in the anterior Hippocampus (HC; Figure 2B; 381–906 ms; $p < 0.01$; cluster-controlled for multiple comparisons using permutation statistics; see STAR Methods). Even though the latter effect in the hippocampus appeared broadband at the group level, patients exhibited clear peaks during TBF trials, most often at 8 Hz (median ± SD peak frequency: 8.5 ± 2.4 Hz, Figures S1B–S1D). Importantly, this analysis was performed after correcting for the unequal number of subsequently remembered and forgotten trials (STAR Methods). The frequency peak of 8 Hz corresponded to the predominant frequency in the hippocampus across all trials and time points (Figures S1F and S1G). The individual peaks were relatively narrowband with a full width at half maximum of 1.5 ± 0.6 Hz (mean ± SD, Figure S1E). As the exact center frequencies of these peaks varied between patients (7–16 Hz, see Figures S1B–S1D), the instruction effect appeared to be much broader in frequency at the group level.

Next, we analyzed whether neural activity after task instructions (i.e., locked to cue onset) was associated with subsequent

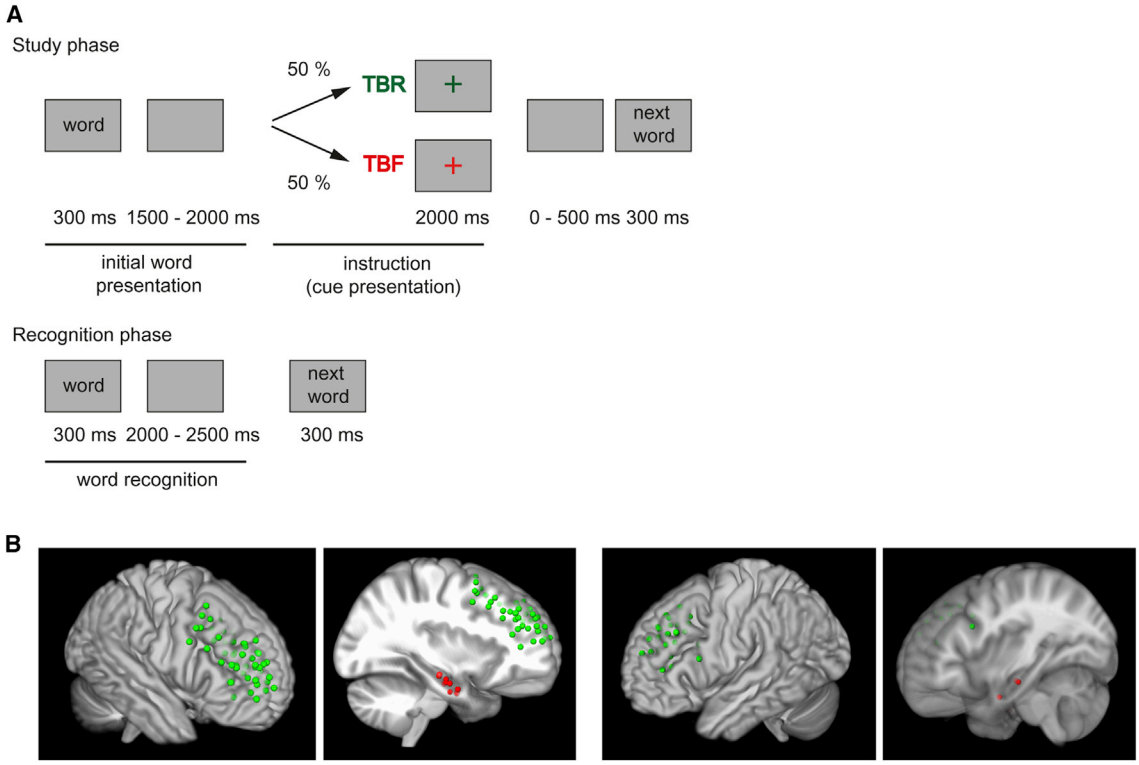

Figure 1. Experimental Design and Electrode Locations

(A) Experimental design (STAR Methods).

(B) Location of the selected DLPFC (green) and hippocampal (red) virtual electrodes of all patients in Montreal Neurological Institute (MNI) space, mapped onto the MNI template. For three patients with hippocampal electrodes (two right, one left hemisphere), a post-implantation MRI was not available. In this case, we selected the two most anterior contacts of the implanted electrodes for analysis. Locations of DLPFC electrodes were visually verified in all individual MRIs in native space.

memory performance. In DLPFC, theta activity was generally higher for subsequently forgotten items, irrespective of the instruction given (Figure 2C; main effect of "memory"; 4–5 Hz, 544–1066 ms; $p = 0.04$). By contrast, in HC, we only observed a significant subsequent memory effect for TBR items (Figure 2D; instruction x memory interaction: 3–4 Hz, 347–1,043 ms; $p = 0.046$). Enhanced low theta power during TBR trials was associated with successful memory formation (Figure 2Di; post hoc cluster analysis of subsequently remembered TBR items (TBR-R) versus subsequently forgotten TBR items (TBR-F): 3–4 Hz, 388–1,203 ms, $p = 0.01$). During TBF trials, theta activity was similarly high for subsequently remembered and forgotten items (Figure 2Dii and Diii; post hoc cluster analysis of subsequently remembered TBF items [TBF-R] versus subsequently forgotten TBF items [TBF-F]: largest cluster: $p > 0.51$; post hoc paired two-tailed t tests on power within the significant time-frequency window of the interaction effect [3–4 Hz, 347–1,043 ms]: TBR-R versus TBR-F: $t_{14} = 5.0$, $p < 0.001$; TBF-R versus TBF-F: $t_{14} = -0.7$, $p = 0.5$; TBR-R versus TBF-R: $t_{14} = 1.2$, $p = 0.24$; TBR-R versus TBF-F: $t_{14} = 0.5$, $p = 0.65$; TBR-F versus TBF-R: $t_{14} = -2.9$, $p = 0.01$; TBR-F versus TBF-F: $t_{14} = -3.8$, $p < 0.01$). On the other hand, a cluster-based permutation analysis of hippocampal power within the predominant oscillatory frequency of the instruction effect (8Hz; see Figures S1B–S1H) revealed subsequent memory effects for TBF trials, but not TBR trials, with significantly higher power for subsequently forgotten than remembered TBF items (Figure 2E; interaction effect: 687–788 ms, $p = 0.049$; main effect instruction: 509–864 ms, $p < 0.01$; main effect subsequent memory: 645–702 ms, $p = 0.07$; post hoc cluster analyses of TBR and TBF items: subsequent memory effect TBF: 8 Hz, 641–805 ms, $p = 0.03$; subsequent memory effect TBR: no clusters found). Additional paired two-tailed t tests between the four conditions in the time-frequency window of the interaction effect (8 Hz, 687–788 ms) confirmed these results (Figure 2F; TBR-R versus TBR-F: $t_{14} = 1.1$, $p = 0.28$; TBR-R versus TBF-R: $t_{14} = -0.98$, $p = 0.34$; TBR-R versus TBF-F: $t_{14} = -3.9$, $p < 0.01$; TBR-F versus TBF-R: $t_{14} = -2.1$, $p = 0.06$; TBR-F versus TBF-F: $t_{14} = -3.7$, $p < 0.01$; TBF-R versus TBF-F: $t_{14} = -2.8$, $p = 0.02$).

Local and inter-regional effects were specific to anterior HC and not found for the posterior HC (Figures S1J and S1K).

Hippocampal-Prefrontal Interactions: Power Correlations

In order to investigate the relationship between power fluctuations in both regions, we conducted four correlation analyses. First, we calculated interregional power correlations in the time-frequency ranges showing increased oscillatory power for TBF versus TBR trials (DLPFC: 3–5 Hz, HC: 6–18 Hz, 300–1,100 ms). We found that these single-trial power values were

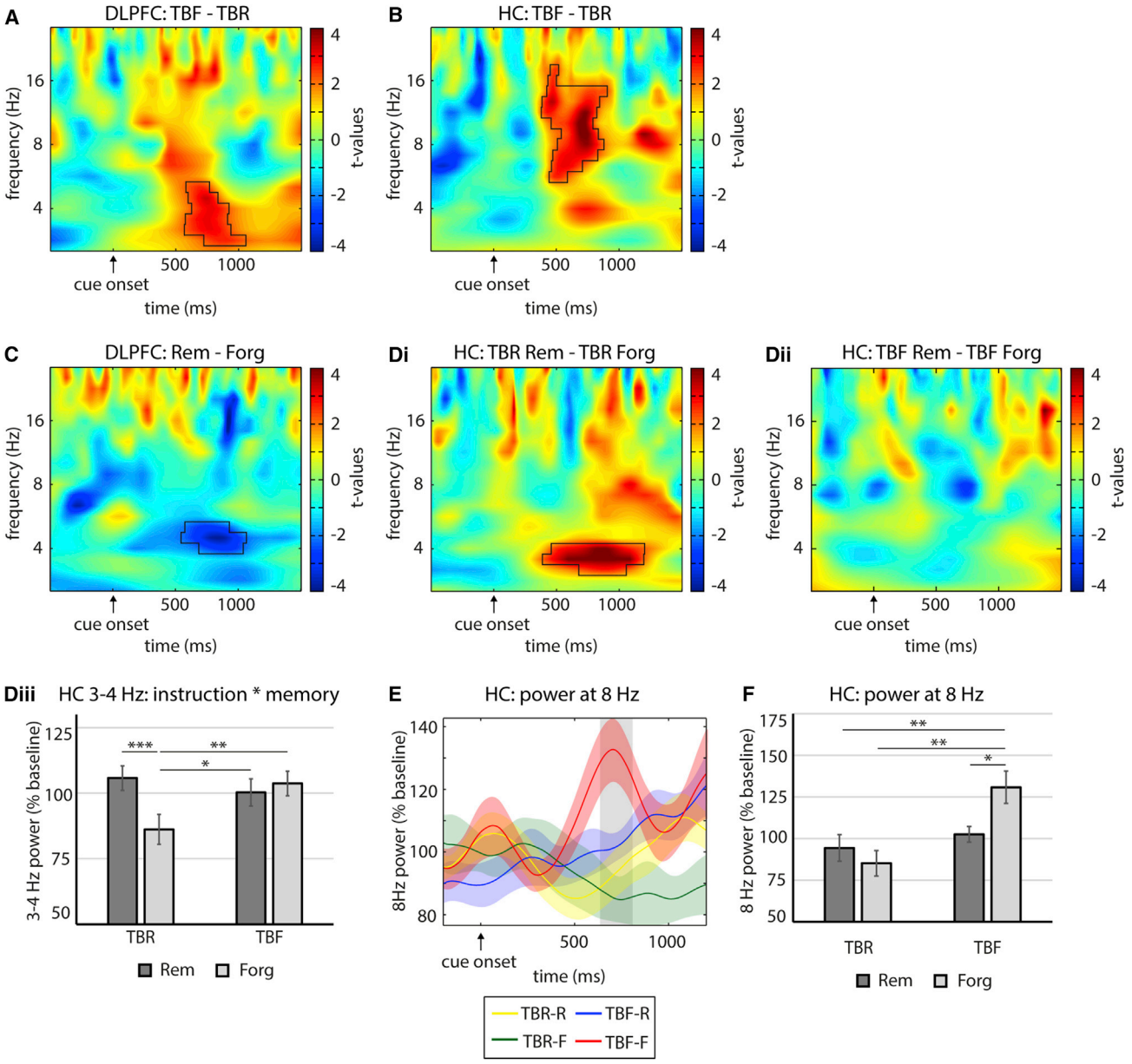

Figure 2. Power Effects in DLPFC and HC

(A–D) Time-frequency resolved comparison of power during two conditions (i.e., t-values; paired-sample t test). The significant clusters are outlined. (A and B) Contrast between power fluctuations during TBF and TBR cues in (A) the DLPFC and (B) the anterior hippocampus. TBF cues were accompanied by enhanced theta oscillations in the DLPFC (3–5 Hz, 568–1,058 ms) and increased theta/alpha/beta power in the hippocampus (6–18 Hz, 381–906 ms). (C and D) Comparison of power values for subsequently remembered and forgotten items. (C) Main effect of subsequent memory in the DLPFC, with reduced theta power for subsequently remembered items (4–5 Hz, 544–1,066ms). (D) Interaction between instruction and subsequent memory in the hippocampus. Follow-up analyses assessed subsequent memory effects separately for TBF and TBR items. (Di) After TBR instructions, subsequent memory was accompanied by enhanced hippocampal theta oscillations (3–4 Hz, 388–1,203 ms). (Dii) Lack of subsequent memory effect after TBF instructions. (Diii) Post-hoc paired two-tailed t tests comparing mean (illustrated ± standard error) theta power during the four conditions (TBR-R, TBR-F, TBF-R, TBF-F) within the significant time-frequency window of the interaction effect (3-4Hz, 347-1043ms). *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

(E) Mean (± standard error) power time-series across patients at the frequency where most patients exhibited a power peak in the HC (8 Hz) (interaction effect: 687–788 ms; subsequent memory effect TBF: 641–805 ms, $p = 0.03$; subsequent memory effect TBR: no clusters found). The significant time period of the contrast between TBF-F and TBF-R items is shaded in gray.

(F) Bar plot illustrating mean (± standard error) hippocampal power at 8 Hz in the significant time window of the interaction effect (687–788 ms). Results of post hoc paired two-tailed t tests between the four conditions are indicated: *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. See also Figure S1.

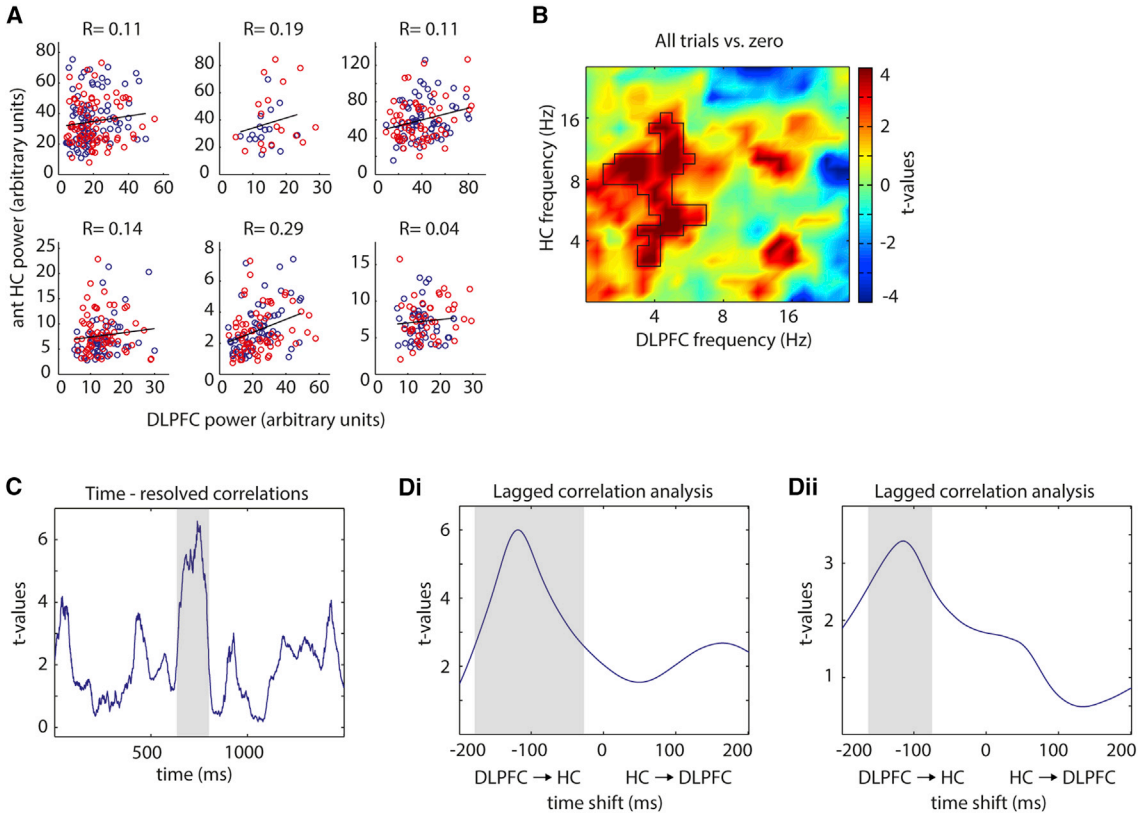

Figure 3. Power Correlations between DLPFC and HC

(A) Single-trial correlations between HC power (6–18 Hz) and DLPFC power (3–5 Hz) in six individual patients with electrodes in both DLPFC and HC. Mean power values were extracted within the time window showing significant power effects (300–1,100ms) for TBR (blue) and TBF (red) trials.

(B) Frequency-resolved correlations across HC/DLPFC frequency pairs in the time window showing significant regional power effects (300–1,100 ms). The significant frequency/frequency cluster is outlined (DLPFC: 3–6 Hz, HC: 3–16 Hz).

(C) Time-resolved correlations between mean HC (6–18 Hz) and DLPFC power (3–5 Hz) at each time point. Significant time periods are shaded in gray (638–796 ms).

(D) Power correlations as a function of time shift of the DLPFC time series in relation to the HC time series.

(Di) Time lags resulting in significant correlations (lagged correlation analysis across all trials);

(Dii) direct comparison of lagged correlations between TBF and TBR trials. Significant time shifts are shaded in gray (all trials versus zero: 27–178 ms; TBF versus TBR: 75–163 ms). See also Figure S2 and S3.

positively correlated across the two regions (Figure 3A; Wilcoxon signed rank test comparing within-subject inter-regional correlation coefficients against zero on the group level: $p = 0.03$). Two additional analyses incorporating all possible low-frequency pairs (2–29 Hz) and time-resolved power correlation values, respectively (see STAR Methods), demonstrated that these power correlations were specific with regard to frequency (Figure 3B; cluster-analysis between 2–29 Hz: DLPFC 3–6 Hz ~HC 3–16 Hz: $p < 0.001$, all other clusters $p > 0.15$) and time (Figure 3C; DLPFC 3–5 Hz and HC 6–18 Hz power values only correlated between 638–796 ms: $p < 0.001$, all other clusters: $p > 0.09$). Power correlations were not different between TBF and TBR trials (Wilcoxon signed rank test comparing within-subject inter-regional correlation coefficients for TBR and TBF trials: $p = 0.3$). Correlations remained significant when we corrected for power differences between TBR and TBF trials (Wilcoxon signed rank test of within-subject inter-regional correlation coefficients for all trials: $p = 0.03$; Figure S3A). Furthermore, there was no difference in inter-regional power correlations for subsequently remembered and forgotten words (Wilcoxon signed rank test: $p = 0.69$).

Finally, we analyzed whether correlations increased when DLPFC and HC time series were shifted in time, i.e., when a temporal lag was introduced between the two time courses (Figure 3D). Indeed, a permutation-based cluster analysis on correlation coefficients as a function of different temporal lags (up to 200 ms in both directions, see STAR Methods) revealed that DLPFC power correlated higher with HC activity during a later time period. Significant correlations were observed at lags between 27 and 178 ms ($p = 0.018$, all other clusters: $p > 0.07$) and peaked at a lag of 118 ms (Figure 3Di). In other words, HC power followed DLPFC power with a time lag of around 120 ms. Interestingly, the lagged correlation values were significantly higher for TBF than TBR trials (Figure 3Dii; 75–163 ms; peak at 115 ms; $p = 0.042$), while the lags at which power correlations were maximal occurred in a similar time range (TBF: 121 ms, TBR: 99 ms; Wilcoxon signed rank test between individual peak correlation time points for TBF and TBR trials: $p = 0.62$).

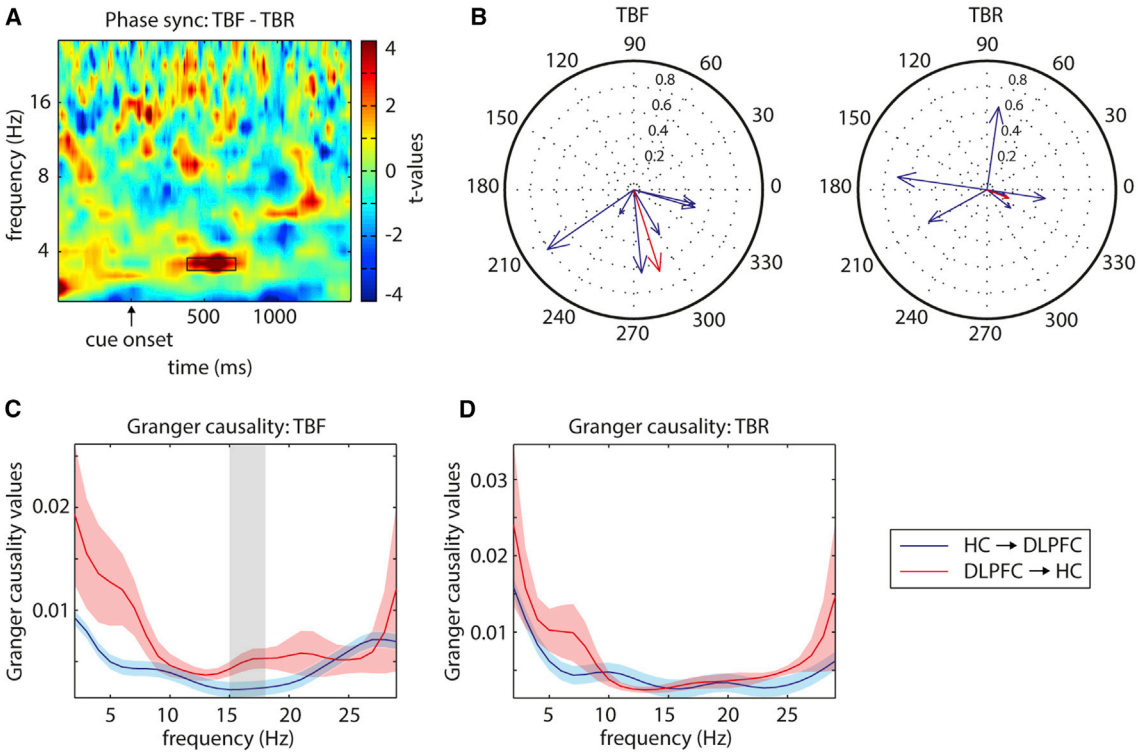

Figure 4. Phase Synchronization and Granger Causality

(A) Time-frequency resolved comparison of phase synchronization between DLPFC and HC during TBF versus TBR trials (paired-sample t test). The significant cluster is outlined (4 Hz, 384–718 ms).

(B) Mean phase shift between DLPFC and HC theta oscillations within the significant time-frequency window (4 Hz, 384–718 ms) across trials, time points and electrodes for each patient (blue) and across patients (red) during TBF (left) and TBR (right) cues.

(C and D) Granger causality analysis (mean ± standard error) of information transfer: HC → DLPFC (blue), DLPFC → HC (red); TBF trials (C), TBR trials (D). Information transfer is significantly asymmetric for TBF, but not TBR trials in the beta band (15–18Hz). Significant frequencies are shaded in gray. See also Figure S3.

There was no effect of subsequent memory on the lagged correlation values (no clusters found; Figure S3B). Very similar results were obtained when outliers were excluded using robust Skipped-Spearman correlations (Figure S2).

Hippocampal-Prefrontal Interactions: Phase Synchronization

We observed enhanced DLPFC-HC phase synchronization in the theta band for TBF compared to TBR trials (Figure 4A; 4Hz, 384–718ms, $p = 0.03$, all other clusters: $p > 0.16$). A non-parametric circular one-sample test comparing phase shifts against zero on a group level [14] showed a significant phase delay between DLPFC and HC signals during TBF trials (Figure 4B; $p = 0.03$, mean across patients: 288°/−72°), but not during TBR trials ($p = 0.69$, mean across patients: 339°/−21°). There was no subsequent memory effect (all clusters: $p > 0.19$) and no interaction between instruction and subsequent memory on phase synchronization values (all clusters: $p > 0.11$). Phase synchronization during processing of TBF items was higher than during baseline (Figure S3C; Wilcoxon signed rank test: TBF PSV versus baseline PSV: $p = 0.03$). In contrast, phase synchronization during TBR trials was lower than during baseline (TBR PSV versus baseline PSV: $p = 0.03$).

Hippocampal-Prefrontal Interactions: Granger Causality

In order to use both power and phase information to investigate the directionality of DLPFC-HC interactions, we calculated frequency-resolved Granger causality in the time-window showing significant power effects (Figures 4C and 4D; 300–1,100 ms). This analysis confirmed the directionality of effects observed in the lagged power correlation analysis. A cluster-based permutation test with the factors "direction" (DLPFC → HC versus HC → DLPFC) and "instruction" (TBR versus TBF) revealed a main effect of direction, with higher DLPFC → HC than HC → DLPFC information flow between 15 and 18Hz ($p = 0.03$), and an interaction between direction and instruction in the same frequency range ($p = 0.017$). There was no main effect of "instruction." A post hoc cluster analysis directly comparing information flow in the two directions after TBF instructions revealed significantly higher DLPFC → HC than HC → DLPFC information flow (Figure 4C; 15–18 Hz, $p < 0.001$). This effect was stable across different model orders ranging from 10–13, corresponding to time lags of 100–130 ms. After TBR instructions, information flow was equally strong in both directions (Figure 4D; TBR: DLPFC → HC versus HC → DLPFC, all clusters $p > 0.1$). The analysis of time-reversed data revealed a flip in directionality of

the interactions between the HC and DLPFC in the beta band (Figure S3D). This indicates that the instruction-dependent asymmetry in Granger causality during TBF instructions was indeed a result of true time-lagged causal interaction between these two regions (see STAR Methods).

Our results indicate that directed forgetting represents an active process associated with enhanced oscillatory activity in both DLPFC and HC, in accordance with previous scalp EEG studies on cognitive control [10–12, 15–17]. Previous studies suggest that PFC-HC interactions occur in the theta frequency range and reflect memory processes [18–22], while top-down attentional control initiated by the PFC occurs at theta and/or beta frequency. Our findings of DLPFC to HC coupling in both of these frequency ranges integrate these effects and suggest that voluntary forgetting recruits both communication channels. We found converging evidence for time lags of DLPFC-HC coupling between 100 and 130 ms. This lag is larger than monosynaptic conduction delays between DLPFC and HC [23] but is consistent with previous findings in rodents reporting lags of 100–150 ms [19, 24] for PFC-HC interactions. Our results are in accord with a framework wherein one brain region sends an oscillatory signal to a receiving brain region which then shows a time lagged response to that signal. This time-lagged response can be seen in both power and phase, in line with a recent computational model [25].

Our study has several limitations. First, direct electrophysiological recordings from the hippocampus are only possible in epilepsy patients, where effects of epilepsy cannot be totally excluded [26]. Second, as electrode placement is clinically determined, iEEG is subject to relatively sparse spatial sampling. Third, the item-method directed forgetting paradigm does not contain a control condition, and it has to be investigated whether our findings, which are related to verbal material, can be generalized to other kinds of memory contents.

Directly recording from the DLPFC and anterior HC of presurgical epilepsy patients, we show that a cue to intentionally forget an item triggers enhancements of DLPFC theta power and anterior HC theta/alpha/beta power. Both phase- and power-based directional connectivity metrics provided convergent evidence for an influence of DLPFC on HC rather than in the opposite direction.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and can be found with this article online at https://doi.org/10.1016/j.cub.2018.07.042.

ACKNOWLEDGMENTS

We would like to thank Anne Do Lam for help with data acquisition, Miriam Bopp for illustrating electrode positions, and Christian E. Elger for patient management. J.F. received support via SFB 1089 (funded by the DFG). N.A. received support via the DFG (SFB 1280/A02, A03, F02; SFB 874/A11, B11; and project AX82/3).

AUTHOR CONTRIBUTIONS

N.A. and J.F. designed the current data analysis strategy. E.L. and T.R. designed the original experiment. E.L. and C.R.O. collected the data. C.R.O. analyzed the data, with contributions by C.B. and under supervision by N.A. and J.F. C.R.O., J.F., C.B., H.K., S.H., and N.A. interpreted the data and drafted the manuscript. All authors discussed the findings and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 21, 2018
Revised: June 22, 2018
Accepted: July 13, 2018
Published: September 6, 2018

REFERENCES

1. Anderson, M.C., Ochsner, K.N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S.W., Glover, G.H., and Gabrieli, J.D.E. (2004). Neural systems underlying the suppression of unwanted memories. Science *303*, 232–235.
2. Depue, B.E., Curran, T., and Banich, M.T. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. Science *317*, 215–219.
3. Wylie, G.R., Foxe, J.J., and Taylor, T.L. (2008). Forgetting as an active process. An FMRI investigation of item-method-directed forgetting. Cerebral cortex (New York, N.Y.: 1991) *18*, 670–682.
4. Hanslmayr, S., Volberg, G., Wimber, M., Oehler, N., Staudigl, T., Hartmann, T., Raabe, M., Greenlee, M.W., and Bäuml, K.-H.T. (2012). Prefrontally driven downregulation of neural synchrony mediates goal-directed forgetting. J. Neurosci. *32*, 14742–14751.
5. Jensen, O., Bonnefond, M., and VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. Trends Cogn. Sci. *16*, 200–206.
6. Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res. Brain Res. Rev. *53*, 63–88.
7. Worden, M.S., Foxe, J.J., Wang, N., and Simpson, G.V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. J. Neurosci. *20*, RC63.
8. Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., Gruber, W.R., and Birbaumer, N. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur. J. Neurosci. *22*, 2917–2926.

9. Hummel, F., Andres, F., Altenmüller, E., Dichgans, J., and Gerloff, C. (2002). Inhibitory control of acquired motor programmes in the human brain. Brain *125*, 404–420.
10. Hanslmayr, S., Pastötter, B., Bäuml, K.-H., Gruber, S., Wimber, M., and Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. J. Cogn. Neurosci. *20*, 215–225.
11. Oehrn, C.R., Hanslmayr, S., Fell, J., Deuker, L., Kremers, N.A., Do Lam, A.T., Elger, C.E., and Axmacher, N. (2014). Neural communication patterns underlying conflict detection, resolution, and adaptation. J. Neurosci. *34*, 10438–10452.
12. Cavanagh, J.F., Cohen, M.X., and Allen, J.J.B. (2009). Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. J. Neurosci. *29*, 98–105.
13. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190.
14. Berens, P. (2009). CircStat. A MATLAB Toolbox for Circular Statistics. J. Stat. Soft *31*.
15. Depue, B.E., Ketz, N., Mollison, M.V., Nyhus, E., Banich, M.T., and Curran, T. (2013). ERPs and neural oscillations during volitional suppression of memory retrieval. J. Cogn. Neurosci. *25*, 1624–1633.
16. Cano, M.E., and Knight, R.T. (2016). Behavioral and EEG Evidence for Auditory Memory Suppression. Front. Hum. Neurosci. *10*, 133.
17. Waldhauser, G.T., Bäuml, K.-H.T., and Hanslmayr, S. (2015). Brain Oscillations Mediate Successful Suppression of Unwanted Memories. Cerebral cortex (New York, N.Y.: 1991) *25*, 4180–4190.
18. Anderson, K.L., Rajagovindan, R., Ghacibeh, G.A., Meador, K.J., and Ding, M. (2010). Theta oscillations mediate interaction between prefrontal cortex and medial temporal lobe in human memory. Cerebral cortex (New York, N.Y.: 1991) *20*, 1604–1612.
19. Siapas, A.G., Lubenov, E.V., and Wilson, M.A. (2005). Prefrontal phase locking to hippocampal theta oscillations. Neuron *46*, 141–151.
20. Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P.L., Gioanni, Y., Battaglia, F.P., and Wiener, S.I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal- prefrontal network upon learning. Neuron *66*, 921–936.
21. Jones, M.W., and Wilson, M.A. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. PLoS Biol. *3*, e402.
22. Gruber, M.J., Hsieh, L.-T., Staresina, B., Elger, C., Fell, J., Axmacher, N., and Ranganath, C. (2017). Theta Phase Synchronization Between The Human Hippocampus And The Prefrontal Cortex Supports Learning Of Unexpected Information. bioRxiv. https://doi.org/10.1101/144634.
23. Jay, T.M., Thierry, A.-M., Wiklund, L., and Glowinski, J. (1992). Excitatory Amino Acid Pathway from the Hippocampus to the Prefrontal Cortex. Contribution of AMPA Receptors in Hippocampo-prefrontal Cortex Transmission. Eur. J. Neurosci. *4*, 1285–1295.
24. Wierzynski, C.M., Lubenov, E.V., Gu, M., and Siapas, A.G. (2009). State-dependent spike-timing relationships between hippocampal and prefrontal circuits during sleep. Neuron *61*, 587–596.
25. Parish, G., Hanslmayr, S., and Bowman, H. (2018). The Sync/deSync Model: How a Synchronized Hippocampus and a Desynchronized Neocortex Code Memories. J. Neurosci. *38*, 3428–3440.
26. Parvizi, J., and Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. Nat. Neurosci. *21*, 474–483.
27. Ludowig, E., Möller, J., Bien, C.G., Münte, T.F., Elger, C.E., and Rosburg, T. (2010). Active suppression in the mediotemporal lobe during directed forgetting. Neurobiol. Learn. Mem. *93*, 352–361.
28. Benoit, R.G., and Anderson, M.C. (2012). Opposing mechanisms support the voluntary forgetting of unwanted memories. Neuron *76*, 450–460.
29. Shirhatti, V., Borthakur, A., and Ray, S. (2016). Effect of Reference Scheme on Power and Phase of the Local Field Potential. Neural Comput. *28*, 882–913.
30. Hamamé, C.M., Vidal, J.R., Perrone-Bertolotti, M., Ossandón, T., Jerbi, K., Kahane, P., Bertrand, O., and Lachaux, J.-P. (2014). Functional selectivity in the human occipitotemporal cortex during natural vision: evidence from combined intracranial EEG and eye-tracking. Neuroimage *95*, 276–286.
31. Spaak, E., Bonnefond, M., Maier, A., Leopold, D.A., and Jensen, O. (2012). Layer-specific entrainment of γ-band neural activity by the α rhythm in monkey visual cortex. Curr. Biol. *22*, 2313–2318.
32. Trongnetrpunya, A., Nandi, B., Kang, D., Kocsis, B., Schroeder, C.E., and Ding, M. (2016). Assessing Granger Causality in Electrophysiological Data: Removing the Adverse Effects of Common Signals via Bipolar Derivations. Front. Syst. Neurosci. *9*, 189.
33. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869.
34. Lachaux, J.P., Rodriguez, E., Martinerie, J., and Varela, F.J. (1999). Measuring phase synchrony in brain signals. Hum. Brain Mapp. *8*, 194–208.
35. Cui, J., Xu, L., Bressler, S.L., Ding, M., and Liang, H. (2008). BSMART: aMatlab/C toolbox for analysis of multichannel neural time series. Neural Netw. *21*, 1094–1104.
36. Ding, M., Bressler, S.L., Yang, W., and Liang, H. (2000). Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. Biol. Cybern. *83*, 35–45.
37. Barrett, A.B., Murphy, M., Bruno, M.-A., Noirhomme, Q., Boly, M., Laureys, S., and Seth, A.K. (2012). Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. PLoS ONE *7*, e29072.
38. Haufe, S., Nikulin, V.V., and Nolte, G. (2012). Alleviating the influence of weak data asymmetries on granger-causal analyses. In Latent Variable Analysis and Signal Separation, F. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, eds. (Springer), pp. 25–33.
39. Bastos, A.M., and Schoffelen, J.-M. (2016). A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. Front. Syst. Neurosci. *9*, 175.
40. Wilcox, R. (2004). Inferences Based on a Skipped Correlation Coefficient. J. Appl. Stat. *31*, 131–143.
41. Pernet, C.R., Wilcox, R., and Rousselet, G.A. (2013). Robust correlation analyses: false positive and power validation using a new open source matlab toolbox. Front. Psychol. *3*, 606.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw and analyzed iEEG data	This paper	Available upon reasonable request
Stimuli	This paper	Available upon reasonable request
Software and Algorithms		
Harmonie EEG software	Stellate Systems	376 Victoria Avenue, Montreal Canada
Presentation	Neurobehavioral Systems	https://www.neurobs.com
FMRIB Software Library	Oxford Centre for Functional MRI of the Brain	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
PyLocator	University of Bonn	www.pylocator.thorstenkranz.de
Brain Vision Analyzer 2	Brain Products	https://www.brainproducts.com
MATLAB R2011a	MathWorks	https://www.mathworks.com
FieldTrip	Donders Institute for Brain, Cognition and Behavior	http://www.fieldtriptoolbox.org

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Nikolai Axmacher (nikolai.axmacher@rub.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patients

We recorded intracranial EEG from 42 patients with pharmacoresistant epilepsy who had been implanted with intracranial electrodes for diagnostic purposes. Depending on the suspected ictal onset zone, patients had been implanted with frontal subdural electrodes, medial temporal depth electrodes, or both. We restricted our analysis on regions free of any morphological changes visible in the MR images and to the hemisphere contralateral to the origin of epileptic activity. After diagnostics had been completed, we excluded patients with bilateral seizure onset and patients in whom the epileptic focus was localized within a unilaterally implanted area of interest (n = 12). Five additional patients were excluded due to extensive epileptic artifacts. Furthermore, we excluded three patients whose recognition accuracy fell two standard deviations below the mean. In total, this resulted in an exclusion of 20 patients, which yielded a final number of 22 patients (11 female; mean age ± standard deviation: 41 ± 12 years). In more detail, a total of 13 patients with prefrontal electrodes was available for analysis (8 right, 5 left) and 15 patients with hippocampal electrodes (11 in the right, 4 in the left hemisphere). Six patients had electrodes in both regions (5 right, 1 left). Data were partially collected within the framework of a previous study [27], which focused on the event-related potential modulations of directed forgetting.

The study was approved by the local medical ethics committee, was in accordance with the latest version of the Declaration of Helsinki, and all patients signed written informed consent.

METHOD DETAILS

Selection of electrodes

Intracranial recordings from frontal regions were obtained using stainless steel subdural strip or grid electrodes. Medial temporal EEG was recorded with multicontact depth electrodes implanted stereotactically along the longitudinal axis of the hippocampus. All data was sampled at 1000Hz, referenced to linked mastoids and band-pass filtered [0.01Hz (6dB/octave) to 300Hz (12dB/octave)] using the digital EPAS system (Schwarzer, Munich, Germany) and Harmonie EEG software (Stellate, Montreal, Canada).

For patients with prefrontal electrodes, we selected all contacts within the DLPFC. To this end, we inspected post-implantation MRIs visually and transferred all contacts into normalized MNI space using the FMRIB Software Library (FSL 5.0, Oxford Centre for Functional MRI of the Brain, University of Oxford) and PyLocator (www.pylocator.thorstenkranz.de, University of Bonn). We converted all MNI coordinates into Talairach space and determined the nearest gray matter for each coordinate (www.talairach.org). Several previous fMRI studies have shown BOLD signal changes in Brodmann areas 9 and 46 (BA 9/46) during voluntary forgetting, in particular in the middle frontal gyrus [1, 2, 4, 28]. We selected all electrodes that were located within the middle frontal gyrus part of Brodmann areas 9/46. This resulted in a range of 1-9 DLPFC contacts per patient (mean ± SD: 4 ± 2 electrodes). Normalization accuracy was sometimes reduced due to poor MRI quality associated with distortions caused by the intracranial electrodes.

However, we visually confirmed that all selected electrodes were located in DLPFC based on the individual (non-normalized) post-implantation MRI.

In a previous intracranial EEG study, we demonstrated directed forgetting effects on event-related potentials primarily in the anterior portion of the hippocampus [27]. For medial temporal electrodes, we inspected post-implantation MRI images and selected the most anterior electrode located within the hippocampus for the analysis.

Experimental paradigm

We used a paradigm identical to the one in a previous intracranial EEG study [27]. Patients completed 4–6 blocks, each consisting of a study and a recognition phase. In each study block, 50 individual words were presented, 25 of which were instructed to be remembered (TBR) and 25 to be forgotten (TBF). Instructions were cued by a green (TBR) or red (TBF) cross appearing within a jittered interval of 1800-2300ms following word presentation (Figure 1A). After each study phase, patients completed a recognition memory test. Patients were presented with all words from the study phase, randomly intermixed with 50 new words. Patients indicated by button press whether a word had been presented previously ("old") or not ("new") - irrespective of the task instruction at study.

Word blocks were balanced regarding word frequency (mean: 65 per 1 million words according to the CELEX lexical database, version 2.5, Baayen, Piepenbrock, & Gulikers, 1995) and word length (range: 4–7 letters). Each word was randomly assigned to the conditions (TBR, TBF, new). Stimuli were presented using Presentation® software (Version 0.71; Neurobehavioral Systems, Albany, CA, USA).

Preprocessing of iEEG data

We first manually rejected artifacts using the Brain Vision Analyzer 2 software (Brain Products, Munich, Germany). Only trials without artifact were later used for further EEG analysis. On average, 24 ± 35 (mean ± SD) trials across all conditions were discarded from each participant resulting in an average of 55 ± 22 TBR-R, 25 ± 16 TBR-F, 47 ± 21 TBF-R and 34 ± 21TBF-F trials (mean ± SD). We re-referenced our data to neighboring electrodes. Previous studies demonstrated that bipolar referencing optimizes estimates of local activity [29–31] and connectivity patterns between brain regions [32]. To this end, we subtracted the activity of neighboring electrodes from each contact and thus created a set of virtual electrodes with bipolar derivations. Each virtual electrode was therefore located in the middle of two physical contacts. Within the hippocampus, we created bipolar derivations between the first and second most anterior electrode located within the hippocampus. Within the prefrontal cortex, we calculated virtual contacts between all available neighbors. On subdural electrode grids, each contact had 2-4 neighbors, on electrode strips 1-2 neighbors. Only virtual electrodes within the DLPFC were used for the subsequent analysis. This resulted in 2-17 virtual DLPFC contacts per patient (mean ± SD: 7 ± 5 electrodes). Further preprocessing and statistical analysis was performed using the FieldTrip toolbox [33] for MATLAB (The Mathworks, Natick, MA, USA) and MATLAB.

We segmented the data from the study phase locked to the onset of the instruction cue (i.e., indication of the TBR or TBF instruction). Then, we divided the data into four conditions, based on subsequent memory of an item (subsequently remembered/forgotten; -R, -F): To-be-remembered words that were subsequently remembered (TBR-R) or forgotten (TBR-F), and to-be-forgotten words that were subsequently remembered (TBF-R) or forgotten (TBF-F). Line noise was removed by means of a discrete Fourier transform.

Calculation of power and phase synchronization

We convolved the preprocessed data with a continuous complex Morlet wavelet with seven cycles. To minimize edge effects occurring after convolution with a wavelet kernel, we segmented the data into relatively large time intervals of 2 s before until 4 s after instruction cue onset and discarded data outside of our time window of interest (which was 0.5 s pre- until 1.5 s post cue onset) after wavelet convolution. From the wavelet-transformed signal, we extracted power values between 2 and 29 Hz in logarithmically increasing steps. For baseline correction, we averaged pre-cue power values (0.5 s before cue onset until cue onset) for each frequency across time and all trials (independent of the instruction given and of subsequent memory). The time series of power values for the respective frequency in each trial was thereafter divided by this subject-specific, but condition-independent value. This baseline correction procedure precluded potential baseline-induced bias due to condition-specific pre-cue effects. In order to assure that our results were not confounded by our choice of baseline, we conducted several control analyses. First, we investigated the leakage of post-stimulus power effects by convolving a Dirac pulse with a seven cycle Morlet wavelet kernel. At the minimum significant frequency (3 Hz; see results section) and therefore the maximal possible leakage of post-stimulus effects into the pre-stimulus window, the power analysis of the Dirac pulse revealed a half width at half maximum of 327ms and thus a leakage outside of the earliest significant window (interaction effect in the hippocampus: 3 - 4Hz, 347 - 1043ms). Thereafter, we calculated power values using an earlier (0.8 s – 0.5 s before cue onset) baseline window and a baseline that did not incorporate any post-cue data (0.5 s before cue onset until cue onset). To this end, we exclusively extracted the raw data from −500ms until cue onset and performed a wavelet convolution. Data was padded by zeroes on either side in order to avoid edge effects. These calculations revealed results in the same time-frequency range as the original analyses.

In order to assess the presence of oscillatory activity during TBF trials in the HC, we conducted several control analyses. For these analyses we performed a wavelet convolution of TBF trials analogous to our original analysis, but used an equally spaced distance between frequencies (0.5 Hz steps from 2 to 29 Hz) in order to be able to estimate peak width for peaks at different frequencies. In a control analysis, logarithmic scaling for the definition of peak frequencies revealed the same predominant frequency across patients

during TBF and across all trials (Figures S1F–S1H). First, we investigated the distribution of frequency peaks during TBF trials in the HC across patients. Importantly, we corrected for an unequal number of subsequently remembered and forgotten TBF trials in order to ensure that the quantification of TBF power values and the selection of peak power was not flawed by a differential number of forgotten and remembered words in these trials. Similarly, we entered the equally weighted averages of all four conditions (TBR-R, TBR-F, TBF-R, TBF-F) into the subsequent cluster analyses investigating interaction and subsequent memory effects. This analysis approach is equal to a 2x2 repeated-measurement ANOVA, in which main effects and interaction effects are orthogonal.

We averaged power data across TBF trials within the significant time window (381 - 906ms) and extracted for each patient the frequency where power was maximal. When multiple peaks were present (n = 7), we extracted the largest one. We excluded patients who did not exhibit a clear peak (n = 2). In a second analysis, we determined the average width of this frequency peak. To this end, we extracted a 2Hz window around the peak frequency and calculated the full width at half maximum for each patient. Finally, we assured that the instruction effect in the HC was not purely driven by evoked power enhancements phase-locked to cue onset. We performed a wavelet convolution of the event-related potential of each patient, i.e., of the averaged time-resolved signal across trials, and subtracted the result of this analysis from the original time-frequency data. Then, we performed a cluster-based permutation analysis analogous to our original analysis (see Figure S1I).

Phase synchronization was calculated using the phase synchronization value [34]. As unequal trial numbers can confound phase synchronization estimates, we randomly selected as many trials from the condition with the higher number of trials as were measured in the condition with the lower number of trials. Results were averaged across 100 random draws [11]. In order to investigate whether phase synchronization was significant for each condition, we compared the phase synchronization values (PSV) during TBR and TBF trials against baseline. To this end, we averaged PSV across time points in the significant time-frequency window (4 Hz, 384-718ms). Analogous to the power analysis, we used a subject-specific, but condition-independent baseline using a pre- stimulus time window from −500ms until instruction onset. Thereafter, TBR and TBF PSV were compared against baseline using a Wilcoxon signed rank test.

Calculation of Granger causality

In order to corroborate the results of the power and phase synchronization analyses, we calculated the directionality of interactions, incorporating phase and amplitude information. We used fieldtrip and the BSMART toolbox to compute frequency-resolved Granger causality within the significant time interval (300-1100ms, [35]). This window size of 800ms is in the range of previously published results using short time windows from 50-2000ms in order to approximate stationarity [36, 37]. To this end, we first subtracted the event-related potential from the data, down-sampled to 100 Hz and thereafter computed sets of multivariate autoregressive coefficients. The optimal model order was estimated for each patient, electrode pair and condition by means of the BIC criterion (range of model orders: 3-15). We selected a model order of 12 in order to capture causality for all subjects within a sufficient time window. This model order corresponds to a time lag of 120ms similar to the lag with maximal correlation values in the analysis of power correlations (118ms, Figure 3Di). Importantly, our results were robust against changes of model order in the range between 10 and 13, corresponding to a time lag of 100 – 130ms. Thereafter, Granger causality was calculated based on the spectral transfer matrices computed from the autoregressive coefficients (2-29 Hz). To assess the directionality of interactions between the HC and DLPFC statistically, we compared the magnitude of Granger coefficients from 2 to 29 Hz for both directions (HC → DLPFC; DLPFC → HC) and for TBF and TBR trials using a cluster-based permutation test. In order to assure that differences in Granger causality values were caused by true time-lagged, causal relationships and not by differences in signal-to-noise ratio, we computed Granger causality after time-reversal of the signals [38, 39]. The underlying rationale is that time-reversal causes a flip in the dominant direction of interaction only in the presence of true, causal asymmetries. To this end, we inversed the temporal order of the raw data and thereafter calculated Granger causality values.

All measures were calculated for each DLPFC and HC electrode (power values) and each possible DLPFC-HC electrode pair (phase synchronization and Granger causality). Thereafter, data across all electrodes within one brain region or across all electrode pairs was collapsed for each patient and entered into statistical analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of behavioral data

We first evaluated the percentage of subsequently remembered items as a function of the instruction given during encoding (paired-sample two-tailed t test: TBR-R versus TBF-R). Thereafter, we assured adequate memory performance by comparing successfully recognized old TBF and TBR words (TBF and TBR hits) to false alarms (paired-sample two-tailed t tests).

Group statistics of electrophysiological data

Statistics for all measures were performed using a non-parametric cluster-based approach comparing sum cluster t-values with surrogates obtained by randomly assigning condition labels to the average data from each participant [13]. To this end, we performed three t tests comparing main effects of instruction (TBR versus TBF) and subsequent memory (remembered versus forgotten) and interaction effects between these two factors (TBR-R – TBR-F versus TBF-R – TBF-F) and compared these to the results of 10,000 permutations. These tests were performed in a frequency spectrum between 2 and 29 Hz and within two time intervals (pre-stimulus interval: 500ms up to instruction cue onset; post stimulus interval: instruction cue onset up to 1500ms). This trial length

(−500ms to 1500ms) was chosen to minimize artifacts in low frequency bands elicited by presentation of the previous (1800-2300ms before cue onset) and next stimulus (2000-2500ms after cue onset).

Inter-regional power correlations

We performed four correlation analyses between power values in both regions. To this end, we used non-baseline corrected single-trial power data in order to preclude any baseline-induced confounding effects on the results. All correlations were calculated across all available trials (independent of instruction and subsequent memory). We excluded trials with power values exceeding the mean power across trials in either region by more than two standard deviations (11 - 25 excluded trials, median: 20 trials). This resulted in a median of 123 remaining trials per patient (range: 33 - 168 trials). In each of the four analyses, Fisher-z-transformed Spearman correlation coefficients between HC and DLPFC power were correlated across trials for each subject. The significance on a group level was subsequently determined by a two-sided Wilcoxon signed rank test testing the null hypothesis that the Fisher-z-transformed Spearman correlation coefficients derived from a distribution with zero median. As this method mainly controls for univariate outliers, we performed the same correlation analyses with a method protecting against bivariate outliers by considering the overall structure of the data [40]. In addition to the Spearman-correlation coefficients from our first analysis, we thus computed Skipped-Spearman correlations [41]. Since the latter method has not been validated for n > 200 according to the authors, we refrained from calculating robust correlations for the temporally lagged time series (300-1100ms, 801 data points).

First, we investigated the inter-regional relationship between power changes in the respective frequencies in the overlapping time window of the respective intra-regional power effects. To this end, we averaged power values between 300 and 1100ms for each patient and each trial within the significant frequency windows in each region (DLPFC 3 - 5 Hz; HC 6 - 18 Hz). Thereafter, we computed Fisher-z-transformed Spearman correlation coefficients between HC and DLPFC power across all trials for each subject. To ensure that potential correlations were not driven by instruction-related power differences, we performed an additional analysis correcting for power differences between conditions. To this end, we subtracted the mean power across all TBR trials (0-1500ms) from each TBR trial. The same was done for TBF trials. Thereafter, we extracted mean power values for all trials, again calculated Fisher-z-transformed Spearman correlation coefficients and performed a two-sided Wilcoxon signed rank test against zero (dependent variable: correlation strength).

Second, we scrutinized the temporal selectivity of these correlations. To this end, we averaged time-resolved power values (0-1500ms) across the significant frequency windows (DLPFC 3-5 Hz, HC 6-18 Hz) for each trial and region. We used 3-cycle wavelets in order to maximize temporal resolution. We then computed Spearman's correlation coefficients across trials for each time point and evaluated significant effects using a cluster-based permutation analysis on the time series of Fisher-z-transformed correlation coefficients testing against zero.

Third, we tested the specificity of power correlations with regard to frequency. To this end, we calculated mean power values for each trial and frequency (2-29Hz logarithmically increasing steps) within the significant time interval (300-1100ms). Thereafter, we computed Spearman's correlation coefficients for all possible frequency pairs (2-29Hz x 2-29Hz). Here, we tested significance on a group level by means of a two-dimensional cluster-based permutation analysis on Fisher-z-transformed correlation coefficients (testing against zero).

Finally, we investigated the directionality of power fluctuations by introducing a varying delay period between the DLPFC and HC power time series. We shifted the DLPFC theta power time series of each trial in time relative to HC theta/alpha/beta power values (−200ms to +200ms in 1ms steps) and calculated Spearman's correlations coefficients across time points within the significant time window (300-1100ms) for each time shift. Thereafter, we again performed a one-factorial cluster-based permutation analysis on the time series of Fisher-z-transformed correlation coefficients testing against zero (dependent variable: correlation strength, independent variable: time shift).

DATA AND SOFTWARE AVAILABILITY

All data, analysis code, and stimuli are available upon reasonable request.