NEUROSCIENCE

Hippocampal theta phases organize the reactivation of large-scale electrophysiological representations during goal-directed navigation

Lukas Kunz[1]*, Liang Wang[2,3,4], Daniel Lachner-Piza[1], Hui Zhang[5], Armin Brandt[1], Matthias Dümpelmann[1], Peter C. Reinacher[6], Volker A. Coenen[6], Dong Chen[7], Wen-Xu Wang[7], Wenjing Zhou[8], Shuli Liang[9], Philip Grewe[10], Christian G. Bien[10], Anne Bierbrauer[5], Tobias Navarro Schröder[11], Andreas Schulze-Bonhage[1], Nikolai Axmacher[5]*

Humans are adept in simultaneously following multiple goals, but the neural mechanisms for maintaining specific goals and distinguishing them from other goals are incompletely understood. For short time scales, working memory studies suggest that multiple mental contents are maintained by theta-coupled reactivation, but evidence for similar mechanisms during complex behaviors such as goal-directed navigation is scarce. We examined intracranial electroencephalography recordings of epilepsy patients performing an object-location memory task in a virtual environment. We report that large-scale electrophysiological representations of objects that cue for specific goal locations are dynamically reactivated during goal-directed navigation. Reactivation of different cue representations occurred at stimulus-specific hippocampal theta phases. Locking to more distinct theta phases predicted better memory performance, identifying hippocampal theta phase coding as a mechanism for separating competing goals. Our findings suggest shared neural mechanisms between working memory and goal-directed navigation and provide new insights into the functions of the hippocampal theta rhythm.

INTRODUCTION

Purposefully and persistently following goals over long spatial and temporal distances is at the heart of human behavior. Achieving a goal often entails navigation to the goal, rendering goal-directed navigation an essential basis of everyday life (*1*). However, the neural basis underlying this complex behavior is incompletely understood in humans.

Phenomenologically, goal-directed navigation exhibits similarities with working memory, because the initially defined goal that shall be achieved has to be maintained throughout the entire navigation period—similar to working memory tasks in which items have to be maintained over a short period of time (*2*). In addition, the current goal has to be protected from interference with other goals that require navigation to different goal locations, a cognitive function that resembles the simultaneous but separate maintenance of different mental contents during working memory. Hence, we hypothesized that neural mechanisms similar to the ones underlying working memory may be recruited to accomplish the complex behavioral capacity of goal-directed navigation.

More specifically, both theoretical and empirical working memory studies suggest that different working memory items are represented via distinct patterns of brain activity that are dynamically reactivated during the maintenance period (*3*, *4*). To keep different working memory items apart from each other and to preserve their serial order during the encoding period, the reactivation of item-specific brain activity patterns during maintenance has been theoretically suggested and empirically reported to occur at different phases of a low-frequency oscillation in the theta/alpha frequency range (*3*, *5*, *6*).

Here, we hypothesized that a similar mechanism of theta-coupled replay [here in the sense of repeated reactivation following (*4*)] holds true for goal-directed navigation, although the exact implementation of this principle might differ. Specifically, following previous studies in rodents, goal-specific representations might be widely distributed across the brain (*7*) with a focus on prefrontal regions that represent future paths (*8*) and spatial goals (*9*, *10*). In addition, the coordinating phases of the low-frequency oscillation may stem from the hippocampal theta rhythm, which dominates the local field potential in both rodents and humans during (virtual) spatial navigation (*11*–*14*) and is therefore a promising candidate for organizing multiple competing goals during goal-directed navigation (*15*).

Hence, in the present study, we hypothesized that periods of goal-directed navigation require dynamic reactivation of the desired object that cues for an associated goal location. To this end, we examined intracranial electroencephalography (iEEG) recordings from epilepsy patients performing an object-location memory task in a virtual environment. We used representational similarity analysis (RSA) to identify large-scale electrophysiological representations of different objects that cued for associated goal locations. We then tracked their dynamic reactivation over time-varying periods of goal-directed navigation and related this dynamic reactivation to the hippocampal theta rhythm, which could be directly observed via hippocampal depth electrodes in a subset of our patients. We found

[1]Epilepsy Center, Medical Center—University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg im Breisgau, Germany. [2]CAS Key Laboratory of Mental Health, Institute of Psychology, Beijing, China. [3]CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, China. [4]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China. [5]Department of Neuropsychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Bochum, Germany. [6]University Medical Center, Stereotactic and Functional Neurosurgery, Freiburg im Breisgau, Germany. [7]School of Systems Science, Beijing Normal University, Beijing, China. [8]Department of Epilepsy Center, Tsinghua University Yuquan Hospital, Beijing, China. [9]Department of Neurosurgery, First Affiliated Hospital of General Hospital of PLA, Beijing, China. [10]Bethel Epilepsy Centre, Krankenhaus Mara, Bielefeld, Germany. [11]Kavli Institute for Systems Neuroscience, Centre for Neural Computation, The Egil and Pauline Braathen and Fred Kavli Centre for Cortical Microcircuits, NTNU, Norwegian University of Science and Technology, Trondheim, Norway.
*Corresponding author. Email: lukas.kunz@uniklinik-freiburg.de (L.K.); nikolai.axmacher@rub.de (N.A.)

that electrophysiological representations of different cues locked to different hippocampal theta phases, suggesting that the hippocampal theta cycle enables the separation of competing goals in a given context. Our results identify similarities between working memory and goal-directed navigation and suggest hippocampal theta phase coding as a neural component underlying goal-directed navigation.

RESULTS

Behavioral data

We examined brain-wide iEEG recordings from $N = 22$ presurgical epilepsy patients performing an object-location memory task in a virtual environment adapted from a previous study (*16*) (Fig. 1A, fig. S1, Materials and Methods, and table S1). Briefly, during an initial learning phase (that was excluded from all analyses), patients were asked to navigate toward eight visible objects and memorize their locations. Subsequently, patients completed variable numbers of retrieval trials, depending on compliance. At the beginning of each retrieval trial, patients were cued with an image of one of the eight objects. Goal-directed navigation occurred after cue presentation while patients approached the remembered location (Fig. 1B). Patients then made a response indicating their decision and received feedback via different emoticons. Afterward, the object appeared in its correct location from where it had to be collected by the patients, allowing further learning. In each trial, spatial memory performance was assessed as the Euclidean distance between the response location and the correct location ("drop error").

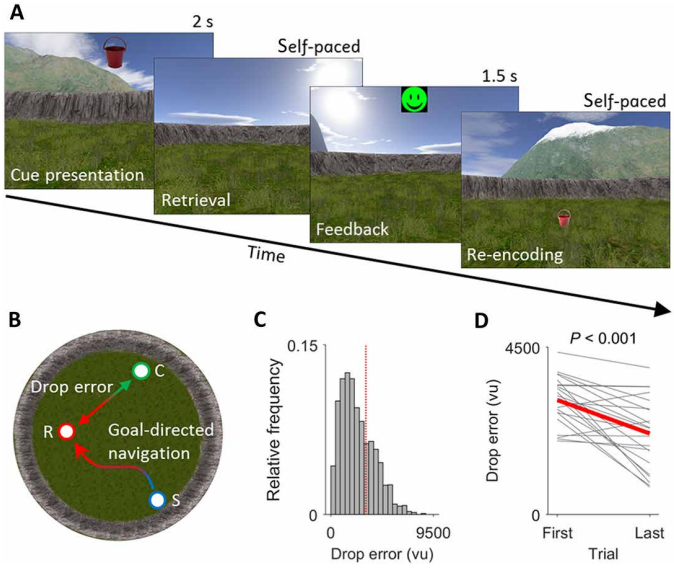

Fig. 1. Virtual navigation task and behavioral data. (**A**) Associative object-location memory task during virtual spatial navigation. At the beginning of the experiment, patients collected eight different objects from eight different locations within the virtual environment. Afterward, patients completed variable numbers of retrieval trials, during which they were first presented with one of the eight objects serving as cue ("cue presentation"). Patients then navigated to the remembered location of that object ("retrieval") and made a response. Following this response, patients received feedback via an emoticon ("feedback") and had to collect the object from its correct location ("re-encoding"). (**B**) Overhead view of the virtual environment (diameter, 9500 vu). Goal-directed navigation occurred after cue presentation, when patients started ("S") navigating to the assumed object location. Starting locations were identical with ending locations from preceding trials and thus varied from trial to trial. The trial-wise drop error was calculated as the Euclidean distance between the response location ("R") and the correct location ("C"). (**C**) Histogram of drop errors across all trials and all patients. Red dashed line, overall chance performance. (**D**) Change in mean drop error across objects between the first and the last trial. Gray lines, patient-wise data; thick red line, average.

Patients completed between 40 and 160 trials (mean ± SD, 102 ± 38 trials) within a time period of 36 to 105 min (mean ± SD, 54 ± 14 min). The mean drop error was 2456.8 ± 709.9 (mean ± SD) virtual units (vu), which is better than patient-wise chance performance (mean ± SD, 3250.2 ± 258.9 vu; paired t test: $t_{21} = -5.22$, $P < 0.001$). Most of the trials (collapsed across patients) showed a drop error better than overall chance performance (70.8%; patient-wise range, 43.0 to 97.6%; Fig. 1C). Furthermore, there was a significant reduction in drop errors (averaged across objects) between the first and the last trial (paired t test: $t_{21} = 4.63$, $P < 0.001$) (Fig. 1D). Retrieval and re-encoding periods had an average duration of 18.6 ± 16.1 s and 14.7 ± 13.1 s (mean ± SD), respectively. The navigation speed was 609 ± 62 vu/s (mean ± SD), on average. These results indicate that patients could successfully perform the task and built reliable associations between the cues and their corresponding goal locations.

Detection of large-scale electrophysiological cue representations

As a precondition for examining the relationship of dynamically reactivated cue representations and hippocampal theta phases during goal-directed navigation, we first had to establish reliable cue representations from the brain-wide iEEG data.

To this end, we used time-resolved spatial RSA (tr-sRSA), which allows the identification of stimulus-specific iEEG patterns in brain-wide activity distributions. That is, we used tr-sRSA to identify a time period during cue presentation for which identical cues exhibited higher neural similarity than different cues (Materials and Methods). Our procedure was grounded in recent advances in decoding approaches for time series data (*17*) and in previous studies using similar types of sRSA (*18*, *19*). Briefly, acquired iEEG data were first low-pass–filtered at 30 Hz, epoched around cue presentation, and converted into independent components (*20*). Our decision to use low-pass–filtered time series data as input to the RSA was motivated by the fact that raw time series data preserve the rich information content of the original signal, including both the power and phase of low-frequency activity (*21*), and have recently been shown to perform well in decoding analyses (*17*). Next, trials were randomly distributed onto two data halves. Within each data half, we calculated one neural vector (NV) across independent components (which is why our RSA approach was labeled "spatial") per cue by averaging across trials of the same cue, separately for each time point within the epoch (which is why our RSA approach was labeled "time-resolved"). Afterward, neural similarity was assessed by calculating the Fisher-z–transformed Spearman correlation coefficient between all combinations of NV_i and NV_j, where i is the cue index of the first data half and j is the cue index of the second data half. Separately for each time point during the epoch, this resulted in an 8×8 confusion matrix of neural similarities between identical (on-diagonal) and different (off-diagonal) cues. Using cluster-based permutation testing (*22*), we then identified a time period during cue presentation, during which identical cues elicited higher neural similarity than different cues, suggesting that this time period contained cue-specific information.

In total, patients contributed recordings from 2330 channels widely distributed across the brain (Fig. 2A). tr-sRSA was applied as

described above to extract neural cue representations. Using cluster-based permutation testing, we found that, during a time period of 256 to 530 ms after cue onset, RSA similarity values were higher for representations of identical cues as compared to different cues (cluster-based permutation testing: $t_{cluster}$ = 799.66, P = 0.019; Fig. 2, B and C). This significant difference allowed us to define neural cue representations whose dynamic reactivation could be examined during subsequent goal-directed navigation (see below). That is, the significant temporal window between 256 and 530 ms after cue onset provided us with a temporal region of interest (tROI) for defining the neural cue representations: Each neural cue representation was obtained by averaging the component-wise iEEG data within the tROI and across trials of the same cue, separately for each patient and each cue.

To corroborate the specificity of identified neural cue representations, we extracted them for both data halves and calculated the similarity of all possible pairs of neural cue representations from the two data halves [following (*23*)], separately for each patient (Materials and Methods). Neural similarity values of identical cues were consistently higher than neural similarity values of different cues (averaged across patients; fig. S2, A and B). Similarly, patient-wise percentage values (assessing how often neural similarity values of identical cues were higher than neural similarity values of different cues) were above 50% chance level for each cue (all t_{21} > 2.28, all P < 0.033). Furthermore, we performed time-resolved spatial multivariate pattern analysis (tr-sMVPA; Materials and Methods) designed to decode the eight different cues from the large-scale electrophysiological patterns. Empirical classifier accuracies were higher as compared to surrogate classifier values during a time period of 402 to 625 ms after cue onset (cluster-based permutation testing: $t_{cluster}$ = 817.70, P = 0.001; fig. S2C). Besides, we observed a strong correlation between classifier accuracy values (empirical minus surrogate;

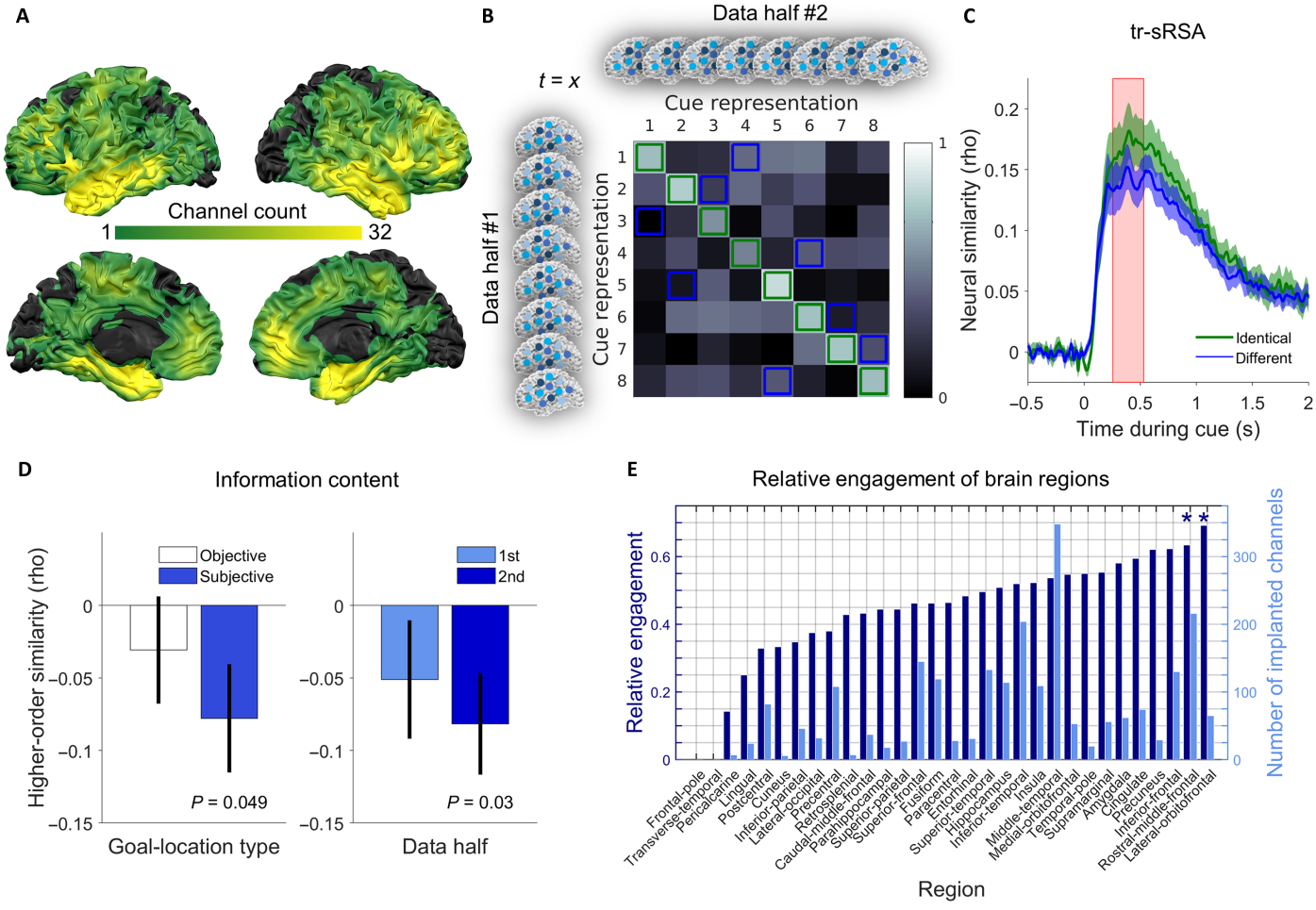

Fig. 2. Identification of large-scale electrophysiological cue representations using tr-sRSA. (**A**) Colored brain surfaces showing the number of channels for each Montreal Neurological Institute (MNI) coordinate. Black coloring, no coverage. (**B**) Analysis principle (illustration). In both data halves, we obtained one cue representation (across channels) per time point ($t = x$) during cue presentation (each small brain, one cue representation). We estimated neural similarity between each pair of cue representations, giving an 8 × 8 confusion matrix. On-diagonal (green squares) and an equal number of off-diagonal values (blue squares; randomly chosen) were extracted, resulting in a time point–specific measure of neural similarity between identical and different cues. (**C**) Higher similarity values for identical as compared to different cue representations between 256 and 530 ms after cue onset (red shaded area). Multivariate iEEG activity contains cue-specific information during this time window, constituting a tROI for subsequent analyses. (**D**) More distinct neural representations of cues with subjectively more similar goal locations (left bar plot), driven by the second data half (right bar plot). Error bars represent SEM. (**E**) Contribution of brain regions to the cue representations ("relative engagement"). Light blue bars, number of implanted channels. *P_{corr} < 0.05 (Bonferroni-corrected for 29 regions).

averaged during the tROI) and neural similarity values (identical minus different; averaged during the tROI) [Spearman's correlation: rho(22) = 0.64, $P = 0.002$], further supporting our conclusion that tr-sRSA enabled the detection of cue-specific neural representations.

Because previous studies revealed stimulus-specific representations based on gamma power [e.g., (*6*, *19*)], we sought to establish a link between the neural cue representations obtained via the filtered raw data (see above) and neural cue representations based on gamma power patterns. To this end, we performed tr-sRSA based on frequency-resolved gamma power (frequencies of 30 to 90 Hz with steps of 4 Hz; Materials and Methods). This analysis revealed a significant cluster ranging from 300 to 410 ms after cue onset within a frequency range of 66 to 82 Hz (cluster-based permutation testing within the tROI: $t_{cluster} = 125.24$, $P = 0.029$; fig. S3). Correlating the patient-wise difference of $RSA_{identical}$ and $RSA_{different}$ of the tr-sRSA based on the raw data (i.e., $RSA_{identical} - RSA_{different}$; averaged during the tROI) with the patient-wise difference of $RSA_{identical}$ and $RSA_{different}$ obtained from the tr-sRSA based on gamma power (i.e., $RSA_{identical} - RSA_{different}$; averaged within the significant cluster) revealed a strong positive correlation [Spearman's correlation: rho(22) = 0.68, $P < 0.001$], demonstrating a link between both types of neural cue representations. This suggests that both types of tr-sRSA capture the activity of widespread neuronal assemblies that exhibit differential responses to the eight different cues.

Inverse relationship between similarity of neural cue representations and similarity of associated subjective goal locations

We next sought to understand the information content of the large-scale electrophysiological cue representations in greater detail. Because patients associated each cue with a unique location in the virtual environment, we hypothesized that the neural cue representations contained spatial information resulting from the associated goal locations. Hence, we analyzed whether pairs of cue representations whose associated goal locations were spatially closer to each other (high similarity of goal locations) exhibited higher or lower neural representational similarity than pairs of cue representations whose associated goal locations were further apart from each other (low similarity of goal locations). We termed the resulting metric "higher-order similarity," because it assessed the correspondence of neural similarity and spatial similarity between pairs of cues and their associated goal locations [fig. S4; see also (*24*)]. Both possibilities (higher versus lower representational similarity) appeared a priori plausible to us: On the one hand, it seems natural that spatially closer goal locations could lead to more similar neural representations of the associated cues (*25*), but on the other hand, recent evidence shows that representations of overlapping routes diverge with learning (*26*, *27*), suggesting inverse relationships between features of the external world and their neural representations. Crucially, the similarity of neural representations may actually be related to assumed ("subjective") rather than objective similarity of goal locations, because neural representations may be more related to their mental content as compared to their corresponding objects in the external world (*25*). Analytically, we thus calculated Spearman correlations between the pairwise similarities of neural cue representations and objective/subjective similarities of goal locations, separately for each patient (Materials and Methods). These analyses revealed that pairs of neural cue representations with subjectively similar goal locations were more distinct from each other than neural cue representations with subjectively dissimilar goal locations (one-sample t test of z-transformed Spearman correlation values against 0: $t_{21} = -2.09$, $P = 0.049$; Fig. 2D). This inverse relationship was particularly present during the second half of the data ($t_{21} = -2.33$, $P = 0.030$; Fig. 2D), suggesting that learning induced the segregation of neural cue representations whose associated goal locations were spatially close to each other, providing further support for the hypothesis that event overlap triggers repulsion of neural representations (*26*, *27*). On a more general level, we also observed that cue representations were more distinct from each other at the end as compared to the beginning of the task (last trial chunk versus first trial chunk, paired t test: $t_{21} = 2.10$, $P = 0.048$; Materials and Methods), which is in line with theoretical accounts suggesting that learning induces refinement of neural representations (*28*).

Following up on this result, we performed several control analyses to detect potential confounding or additional factors determining the information content of the neural cue representations (Materials and Methods). First, we assessed whether basic behavioral characteristics (direction, speed, and acceleration) associated with each cue could account for the pairwise similarity of neural cue representations. We found that none of these three metrics showed a consistent pattern of higher-order similarity: Spearman correlations between the pairwise similarity of neural cue representations and the pairwise similarity of cue-specific direction, speed, or acceleration patterns were not consistently above or below zero (one-sample t tests: all $t_{21} < 1.11$, all $P > 0.282$). Assessing visual similarity of the cue objects via a deep neural net (DNN) did not reveal a consistent relationship between the pairwise similarity of neural cue representations and the pairwise visual similarity of the cue representations at any DNN layer either (all $t_{21} < 2.12$, all $P_{corr} > 0.366$, Bonferroni-corrected for eight DNN layers). Furthermore, specific reward patterns associated with each cue over the course of the experiment only accounted for a small (nonsignificant) amount of the similarity relationships between neural cue representations (one-sample t test: $t_{21} = 1.96$, $P = 0.064$). In summary, these results reveal that the neural cue representations preferentially contained information about the subjective spatial relationship between cue-associated goal locations.

Large-scale electrophysiological cue representations particularly rely on prefrontal regions

To elucidate which brain regions contribute to the cue representations, we performed a jackknife resampling procedure. Specifically, each of the patients' channels was omitted once from the tr-sRSA procedure. For each channel, we then tested whether the area under the curve (AUC) of the tr-sRSA result for identical cues (Fig. 2C, green line) increased or decreased during the tROI when omitting this channel, reflecting a negative ($C_{on} < 0$) or positive ($C_{on} > 0$) contribution of a given channel, respectively. Collapsing across patients, we found that 1200 of the 2330 channels (51.5%) contributed positively to the RSA results (i.e., they increased the neural similarity of identical cues from both data halves), roughly comparable to previous findings (*19*). Of all regions, lateral orbitofrontal cortex ($n = 11$ implanted patients with a total of 65 channels) and rostral middle frontal gyrus ($n = 12$ implanted patients with a total of 216 channels) showed the highest percentages of positively contributing channels (69.8 and 63.4%, respectively). Only the contribution of these two regions exceeded chance level that was estimated by shuffling the channel-wise C_{on} values with respect to the channel labels multiple times, each time calculating the percentage of positively contributing channels for a given region (permutation test, both $P_{corr} < 0.05$,

Bonferroni-corrected for 29 regions; Fig. 2E and Materials and Methods). This result is in accordance with the relevance of these areas for representing goals in rodents and monkeys (*9*, *10*, *29*). Additional analyses showed, however, that the information provided by these prefrontal areas was neither necessary nor sufficient for cue-specific neural representations: Reperforming the tr-sRSA without any channel located in these two regions still revealed higher similarity values for representations of identical cues as compared to different cues within a time period of 305 to 525 ms after cue onset (cluster-based permutation testing: $t_{cluster} = 644.00$, $P = 0.027$). Reversely, performing tr-sRSA based only on activity from these channels did not provide significant cue-specific information (cluster-based permutation testing, $P = 0.400$). Furthermore, iteratively adding channels to those located in lateral orbitofrontal cortex and rostral middle frontal gyrus lead to a clear increase in the difference between $S_{identical}$ and $S_{different}$ during the tROI. Together, these results demonstrate that neural cue representations relied on large-scale electrophysiological patterns and not exclusively on lateral orbitofrontal cortex and rostral middle frontal gyrus (fig. S5).

In addition to analyzing which brain regions increased the neural similarity of identical cues, we also examined which brain regions decreased the neural similarity of different cues from both data halves, thus also contributing to a significant difference between the neural similarity of identical and different cues. Again, we performed a jackknife resampling procedure in which each of the channels was omitted once from the tr-sRSA procedure. For each channel, we then tested whether the AUC of the tr-sRSA result for different cues (Fig. 2C, blue line) increased or decreased during the tROI when omitting this channel, reflecting a negative ($C_{off} < 0$) or positive ($C_{off} > 0$) contribution to the neural similarity of different cues. For each brain region, we then calculated the percentage of positively contributing channels ($C_{off} > 0$), where a low percentage value means that a given brain region decreases the similarity of different cue representations and is thus beneficial for our tr-sRSA results. Of all regions, lateral orbitofrontal cortex, middle temporal gyrus, medial orbitofrontal cortex, and lingual gyrus exceeded chance level (permutation test, all $P_{unc.} < 0.05$; however, no region survived Bonferroni correction for 29 regions; fig. S6).

Furthermore, we computed second-level statistics across all patients examining which brain regions simultaneously increased the neural similarity of identical cues and decreased the neural similarity of different cues (i.e., exhibiting a positive difference between C_{on} and C_{off} values). This analysis revealed a prefrontal cluster extending into rostral middle frontal gyrus (cluster-based permutation testing, $P = 0.038$; fig. S7). In summary, these findings demonstrate that neural cue representations relied on large-scale electrophysiological signals with a focus on prefrontal regions.

Human hippocampal theta oscillations during goal-directed navigation

Since our main goal was to examine the relationship of dynamically reactivated cue representations and the hippocampal theta rhythm during goal-directed navigation, we next sought to characterize hippocampal activity during periods of goal-directed navigation. Because not all patients were implanted with hippocampal channels, analyses focusing on the hippocampus were restricted to a subset of $n = 16$ patients. In each of these 16 patients, we selected one hippocampal channel that showed the clearest event-related potential (ERP) during cue presentation (Materials and Methods, table S2, and Fig. 3A). The average ERP during cue presentation from these channels peaked at 458.1 ± 56.4 ms (mean ± SEM; Fig. 3B). Note that this selection procedure is orthogonal to any analysis focusing on the period of goal-directed navigation after cue presentation.

Theta oscillations in humans during virtual navigation were previously shown to oscillate at a lower frequency than during real-world navigation in rodents and were also shown to occur in bursts rather than continuously during movement (*11*–*14*). To account for these characteristics of human theta oscillations, we applied a recently developed algorithm [termed "MODAL"; for a detailed description, see (*30*)] to the hippocampal recordings of our remaining 16 patients to (i) identify the prevailing theta frequency in our virtual navigation task and (ii) define periods of goal-directed navigation when oscillatory activity in the theta frequency range was present (for three examples, see Fig. 3C). MODAL thus outputs, for a given data segment, frequency bands whose power exceeds the background spectrum during specific time periods. Accordingly, we applied

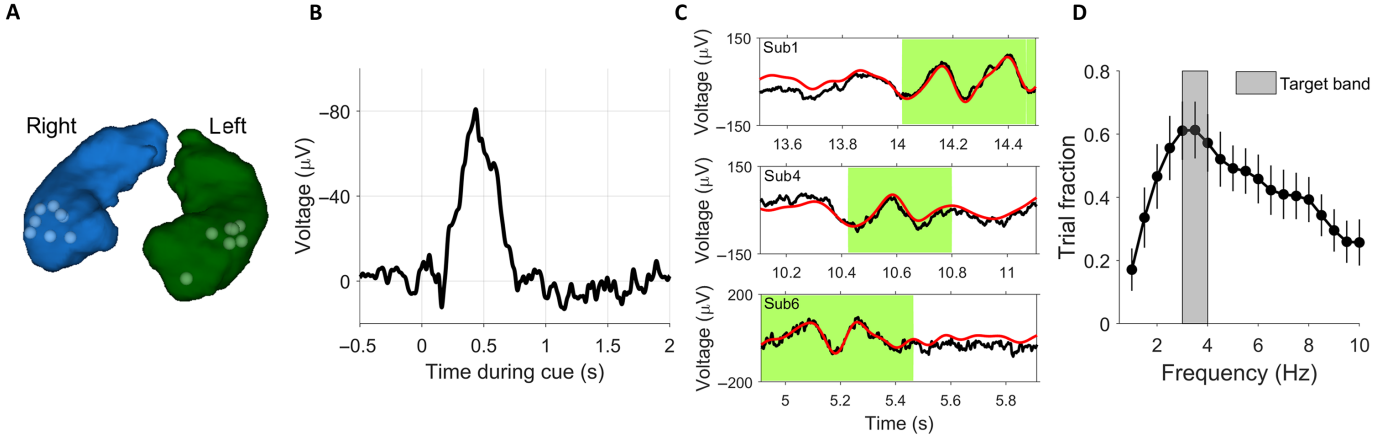

Fig. 3. Hippocampal theta oscillations during goal-directed navigation. (**A**) Depiction of selected hippocampal electrode channels, which were located in the anterior hippocampus. Each white dot represents one channel from a separate patient ($n = 16$). (**B**) Hippocampal ERP during cue presentation. (**C**) Exemplary time periods with theta oscillations during goal-directed navigation from different patients. Black, raw signal; red, low-frequency component of the raw signal (passband, 1 to 10 Hz); green shading, time periods with theta oscillations as detected by MODAL (see Materials and Methods). (**D**) Summary of frequency bands detected by MODAL, across patients, showing that they preferentially occurred at a frequency of 3 to 4 Hz. Dots represent mean, and vertical lines represent SEM.

MODAL to each trial of each patient and kept the frequency band that fell within a frequency range of 1 to 10 Hz (*11*). Across patients, we then calculated the percentage of trials in which a given frequency was contained in the extracted band. This revealed a prevailing theta frequency of 3 to 4 Hz, strongly resembling previous findings (*13*) (Fig. 3D). Instantaneous theta frequencies did not vary as a function of retrieved cue [patient-wise one-way analysis of variance (ANOVA) across cues: all $F < 1.63$, all $P > 0.132$; Materials and Methods]. In addition, MODAL allowed us to define periods of goal-directed navigation with theta oscillations that were then recruited for our analysis of dynamic reactivation of large-scale electrophysiological cue representations at specific hippocampal theta phases (see below). As in previous studies [e.g., (*14*)], these periods occurred in bursts with a mean duration of 364 ± 36 ms (mean ± SEM). On average, 62.6 ± 3.2% (mean ± SEM) of the retrieval periods contained theta oscillations as defined by MODAL.

Dynamic reactivation of large-scale electrophysiological cue representations at specific hippocampal theta phases

We were now in the position to examine the dynamic reactivation of large-scale electrophysiological cue representations and their relationship to hippocampal theta phases during goal-directed navigation. To this end, we first extracted one prototypal neural representation for each of the eight cues by averaging across time within the tROI and across trials of the same cue, separately for each patient. Thus, each prototypal neural cue representation consisted of a $c \times 1$ brain-wide spatial pattern of voltage values, where c refers to the number of independent components in a given patient (no independent components were excluded before this analysis; see Fig. 4A for a schematic illustration of a prototypal neural cue representation).

For a given trial, we then slid the prototypal cue representation whose associated goal location had to be retrieved during this trial across all instantaneous NVs of the retrieval period (Fig. 4A). NVs are the brain-wide iEEG patterns at time point t during retrieval [after converting the channel-wise time series data into independent components using the independent component analysis (ICA) unmixing matrix described in the "Detection of large-scale electrophysiological cue representations" section]. This procedure resulted in a "sliding RSA" time course of similarity values depicting dynamic reactivation of the cue representation over the course of goal-directed navigation (Fig. 4B). A high sliding RSA value means that the corresponding NV (NV_t)—i.e., the brain-wide iEEG activity pattern at time point t during the retrieval period—resembles the cue representation extracted from the tROI during cue presentation. As can be seen from the example in Fig. 4B, sliding RSA values fluctuated between states of high and low resemblance.

The dynamically changing similarity values were then related to the concurrent hippocampal theta rhythm to determine the preferred

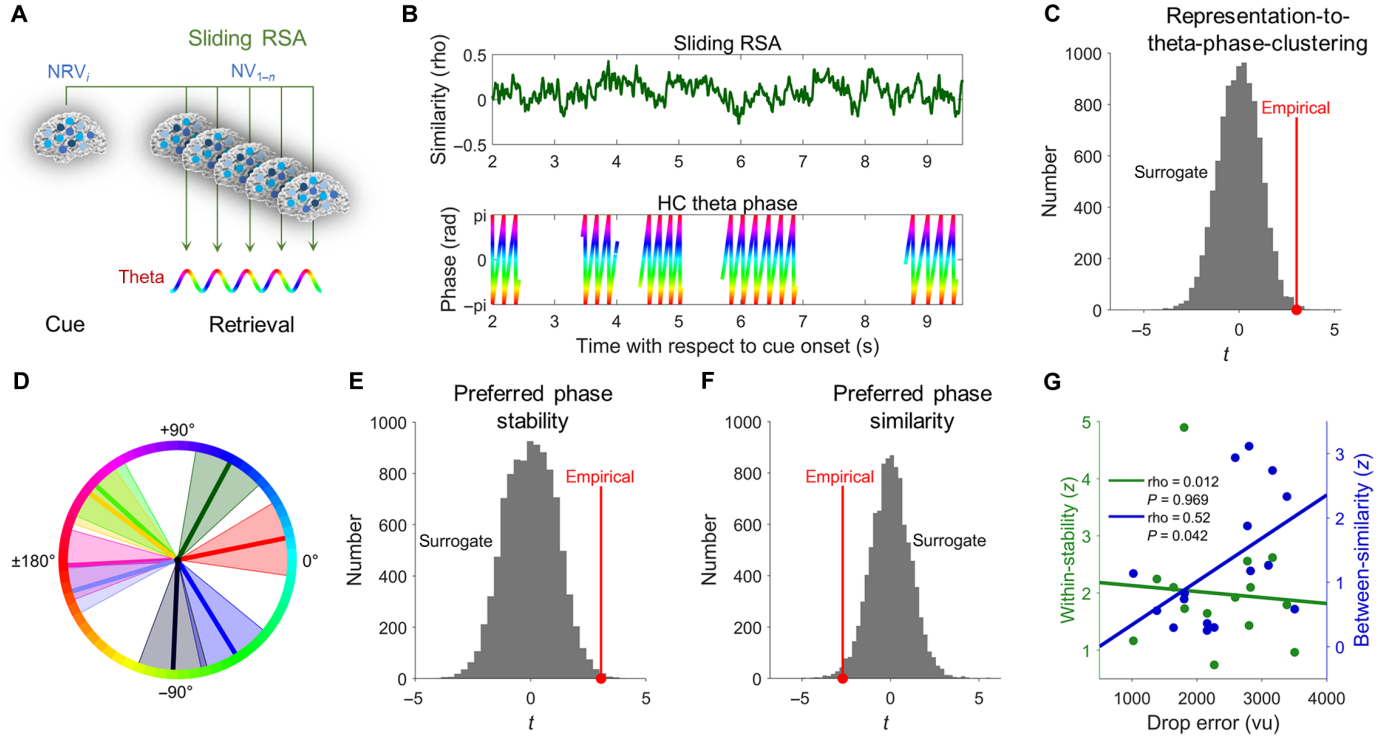

Fig. 4. Dynamic reactivation of cue representations at distinct hippocampal theta phases during goal-directed navigation. (**A**) Analysis procedure of representation-to-theta-phase-clustering. For each cue, we extracted one neural representation vector (NRV_i) within the tROI during cue presentation. For each trial, we then calculated the dynamically changing similarity between NRV_i (i.e., the NRV of cue i whose goal location had to be retrieved during this trial) and all NVs (NV_{1-n}) during the retrieval period. (**B**) This resulted in a sliding RSA time course (top subplot). In addition, we obtained one hippocampal (HC) theta phase for each time point during retrieval where power exceeded the background 1/f spectrum (bottom subplot). For each trial, the preferred theta phase of the sliding RSA values was extracted via the nonparametric Moore-Rayleigh test. (**C**) We assessed the significance of representation-to-theta-phase-clustering by comparing against a surrogate distribution that was obtained by circularly shifting the sliding RSA values against the hippocampal theta phases. Red dot and red line, empirical mean. (**D**) Preferred theta phases of one patient. Bold lines, preferred theta phases; shaded areas, circular SEM. (**E**) Stability of preferred theta phases across trials (within-stability). (**F**) Distinctiveness of preferred theta phases (between-similarity). (**G**) Lower between-similarity values are associated with better spatial memory performance.

theta phase of a given trial. In detail, we estimated the preferred theta phase of an individual trial via the Moore-Rayleigh test, which is a nonparametric extension of the Rayleigh test (*31*) that weights the input phases by a ranked factor (i.e., the sliding RSA values in our case), resulting in the preferred phase and a strength estimation r^* of our "representation-to-theta-phase-clustering." To confirm that we captured representation-to-theta-phase-clustering, we compared the empirical r^* value to surrogate r^* values created by circularly shifting the dynamically changing similarity values with respect to the concurrent theta phases (permutation test, $P = 0.004$; Fig. 4C). For each patient (for an example, see Fig. 4D), we evaluated whether the preferred theta phases of individual cue representations were stable across trials ("within-stability"). We also tested whether the preferred phases of different cue representations were similar to each other ("between-similarity"). We found that individual cue representations clustered at specific theta phases (permutation test, $P = 0.003$; Fig. 4E) and different cue representations clustered at different theta phases (permutation test, $P = 0.012$; Fig. 4F), suggesting that the hippocampal theta cycle provides a means to coherently represent different (competing) goals via phase coding.

Next, we examined the functional relevance of hippocampal theta phase coding for spatial memory performance. We found that lower Rayleigh's $z_{\text{between-similarity}}$ values were associated with lower mean drop errors [Spearman's correlation: rho(16) = 0.52, $P = 0.042$; nonparametric partial correlation controlling for the number of trials: rho(16) = 0.53, $P = 0.044$; nonparametric partial correlation controlling for subject-specific average movement speed: rho(16) = 0.53, $P = 0.043$; Fig. 4G], indicating that patients performed better when the preferred theta phases of the eight cue representations were more distinct from each other. By contrast, $z_{\text{within-stability}}$ values were not related to mean drop errors [Spearman's correlation: rho(16) = 0.01, $P = 0.969$; nonparametric partial correlation controlling for the number of trials: rho(16) = −0.01, $P = 0.999$].

The finding of representation-to-theta-phase-clustering is illustrated in depth for eight trials of one patient (Fig. 5), showing that the eight different cue representations are reactivated at different hippocampal theta phases. The separate reactivation of the eight different cues and their associated goal locations is thus achieved via clustering to different phase ranges of the theta cycle.

As a control, we also analyzed the dynamic reactivation of large-scale electrophysiological cue representations at specific hippocampal theta phases using a subset of $n = 12$ patients. Here, four patients were excluded because of relatively higher levels of epileptic activity in their hippocampal channels (Materials and Methods). This subset of patients showed qualitatively identical results: Representation-to-theta-phase-clustering was significant (permutation test, $P = 0.017$), preferred theta phases of individual cue representations were stable across trials (permutation test, $P = 0.001$), different cue representations clustered at different theta phases (permutation test, $P = 0.027$), and there was a significant relationship between $z_{\text{between-similarity}}$ values and mean drop errors [Spearman's correlation: rho(12) = 0.71, $P = 0.012$; nonparametric partial correlation controlling for the number of trials: rho(12) = 0.65, $P = 0.031$] but not between $z_{\text{within-stability}}$ values and mean drop errors [Spearman's correlation: rho(12) = 0.24, $P = 0.457$; nonparametric partial correlation controlling for the number of trials: rho(12) = 0.21, $P = 0.543$].

In favor of the specificity of this finding, representation-to-theta-phase-clustering was not observed when analyzing time periods during which MODAL did not detect hippocampal theta oscillations (permutation test, $P = 0.414$). Furthermore, representation-to-theta-phase-clustering was not observed at 3 to 4 Hz in the entorhinal cortex as a control region (permutation test, $P = 0.123$; subset of $n = 10$ patients with electrode channels in entorhinal cortex).

Dynamic reactivation of large-scale electrophysiological cue representations at specific prefrontal theta phases

Because previous work in rodents revealed a close relationship between hippocampal and prefrontal theta oscillations (*32*), we finally aimed at extending our findings of a dynamic reactivation of large-scale electrophysiological cue representations at specific hippocampal theta phases to the prefrontal theta rhythm. To do so, we first selected one prefrontal channel in each patient who was implanted in the prefrontal cortex ($n = 13$) (Fig. 6A and Materials and Methods). Parallel to hippocampal channels, exemplary time periods with theta oscillations as defined by MODAL (*30*) are shown in Fig. 6B. Prefrontal theta oscillations occurred in significant temporal association with hippocampal theta oscillations, as revealed by an increased percentage of time in which prefrontal theta oscillations occurred when hippocampal theta oscillations were present as compared to when they were not present (subset of $n = 8$ patients with both hippocampal and prefrontal channels; 67.1% versus 62.2%; paired t test: $t_7 = 2.53$, $P = 0.039$; Fig. 6C). Furthermore, we found increased 3.5-Hz phase coupling (*33*) between prefrontal cortex and hippocampus for good as compared to bad performance trials at the end of goal-directed navigation when decision-making load presumably was highest (last 1.5 s of the retrieval period; paired t test: $t_7 = 5.36$, $P = 0.019$, Bonferroni-corrected for 18 frequencies; Materials and Methods and fig. S8). This result is in line with findings in rodents showing enhanced coupling between hippocampus and prefrontal cortex during task periods associated with peak mnemonic and decision-making load (*32*). Control analyses showed that phase locking values (PLVs) at 3.5 Hz during good performance trials were also significantly higher than surrogate PLVs during good performance trials (paired t test: $t_7 = 2.70$, $P = 0.030$), whereas PLVs at 3.5 Hz during bad performance trials were not higher than surrogate PLVs during bad performance trials (paired t test: $t_7 = -1.95$, $P = 0.092$).

To reveal a dynamic reactivation of large-scale electrophysiological cue representations at specific prefrontal theta phases, we proceeded as described above for the hippocampal theta rhythm. Using MODAL (*30*), we first determined the prevailing theta frequency in the selected prefrontal channels, which showed a peak at 5.5 Hz (Fig. 6D), being significantly faster than in the hippocampus (paired t test: $t_7 = -2.42$, $P = 0.046$). Hence, when subsequently analyzing representation-to-theta-phase-clustering, sliding RSA values were associated with prefrontal theta phases at 5 to 6 Hz during time periods in which MODAL detected prefrontal theta oscillations. We found significant representation-to-theta-phase-clustering (permutation test, $P = 0.003$; Fig. 6E); within-stability values were significantly higher than surrogates (permutation test, $P = 0.006$; Fig. 6F), and between-similarity values were significantly lower than surrogate between-similarity values (permutation test, $P = 0.031$; Fig. 6G). Again, we found a positive relationship between between-similarity values and mean drop errors, meaning that patients with more distinct prefrontal theta phases performed better [Spearman's correlation: rho(13) = 0.63, $P = 0.025$; Fig. 6H; nonparametric partial correlation controlling for the number of trials: rho(13) = 0.73, $P = 0.007$]. Representation-to-theta-phase-clustering was not observed at 3 to 4 Hz (permutation

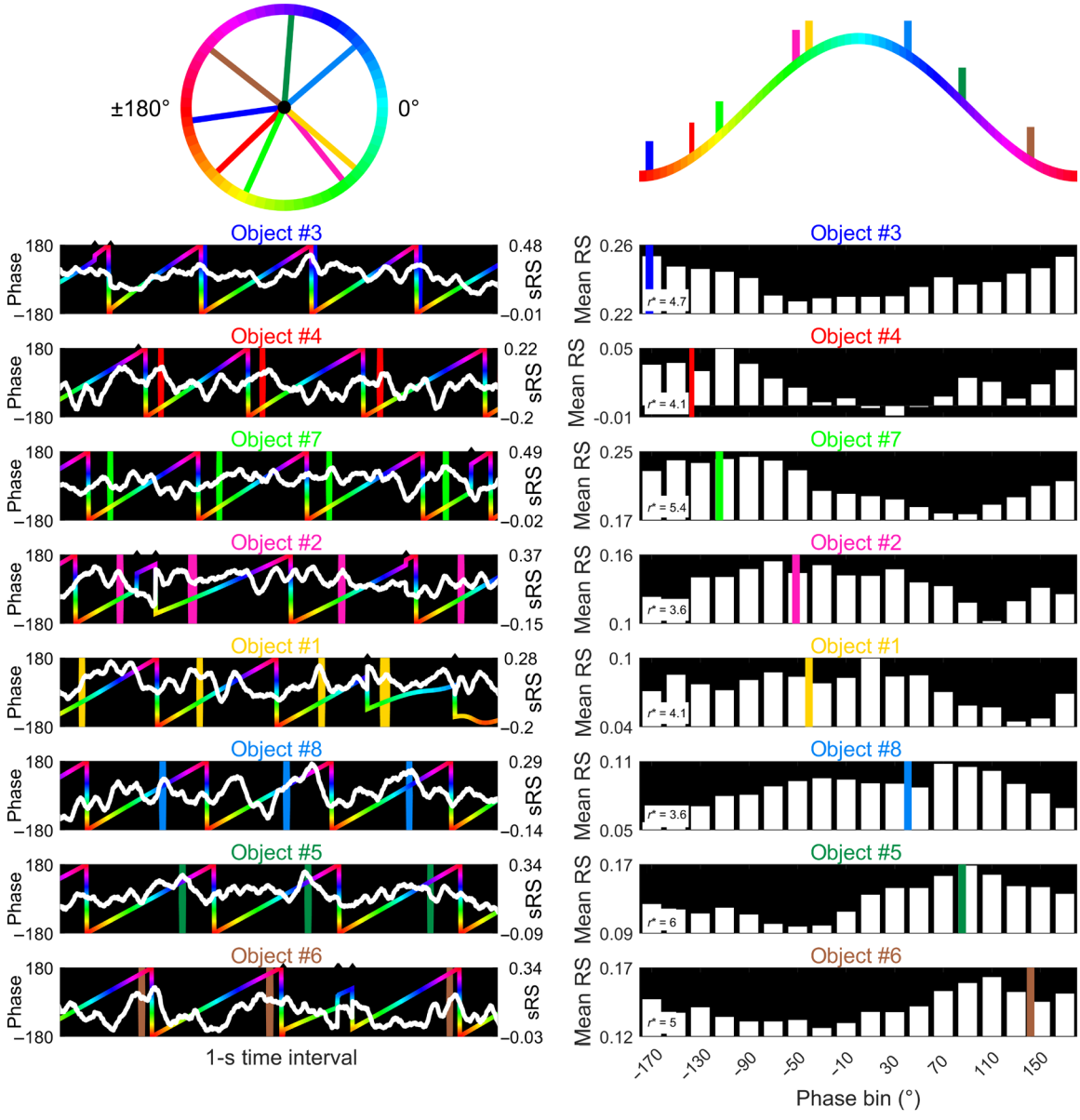

Fig. 5. Exemplary trials of one patient depicting representation-to-theta-phase-clustering. The circle on the top left depicts the preferred theta phases from eight different trials (one for each cue representation). This circle is unfolded on the top right. Color coding corresponds to Fig. 4D. The eight subplots on the left side show sliding RSA time courses (white) and concurrent hippocampal theta phases (colored) during a 1-s time interval of each of the eight trials (corresponding to three to four theta cycles). Shaded areas indicate the time periods during which the phase bin of the preferred theta phase was present. Phase jumps (indicated by small black triangles above each subplot) occur because only periods were included when theta oscillations above the background 1/*f* spectrum were identified by MODAL (these periods were concatenated). The eight right-hand subplots depict the averaged sliding RSA values (separately for each phase bin) from the entire trial. Colored bold lines depict preferred theta phases. *r** values obtained from the Moore-Rayleigh test indicate the strength of representation-to-theta-phase-clustering for the given trial. The eight subplots on the right and left side were ordered according to the preferred theta phase of the dynamically reactivating cue representations. sRS, sliding representational similarity; RS, representational similarity.

test, $P = 0.071$) and was not observed when analyzing time periods during which MODAL did not detect prefrontal theta oscillations (permutation test, $P = 0.299$).

DISCUSSION

The main finding of the current study is that large-scale electrophysiological cue representations dynamically reactivate at specific hippocampal theta phases during goal-directed navigation. In the following, we will discuss our findings with respect to (i) the broader role of hippocampal theta during spatial navigation, (ii) the relationship between our findings and previous working memory studies, and (iii) the function of the prefrontal cortex during goal-directed navigation as revealed in previous rodent studies.

Theta oscillations have been extensively studied during real-world and virtual spatial navigation in both rodents and humans.

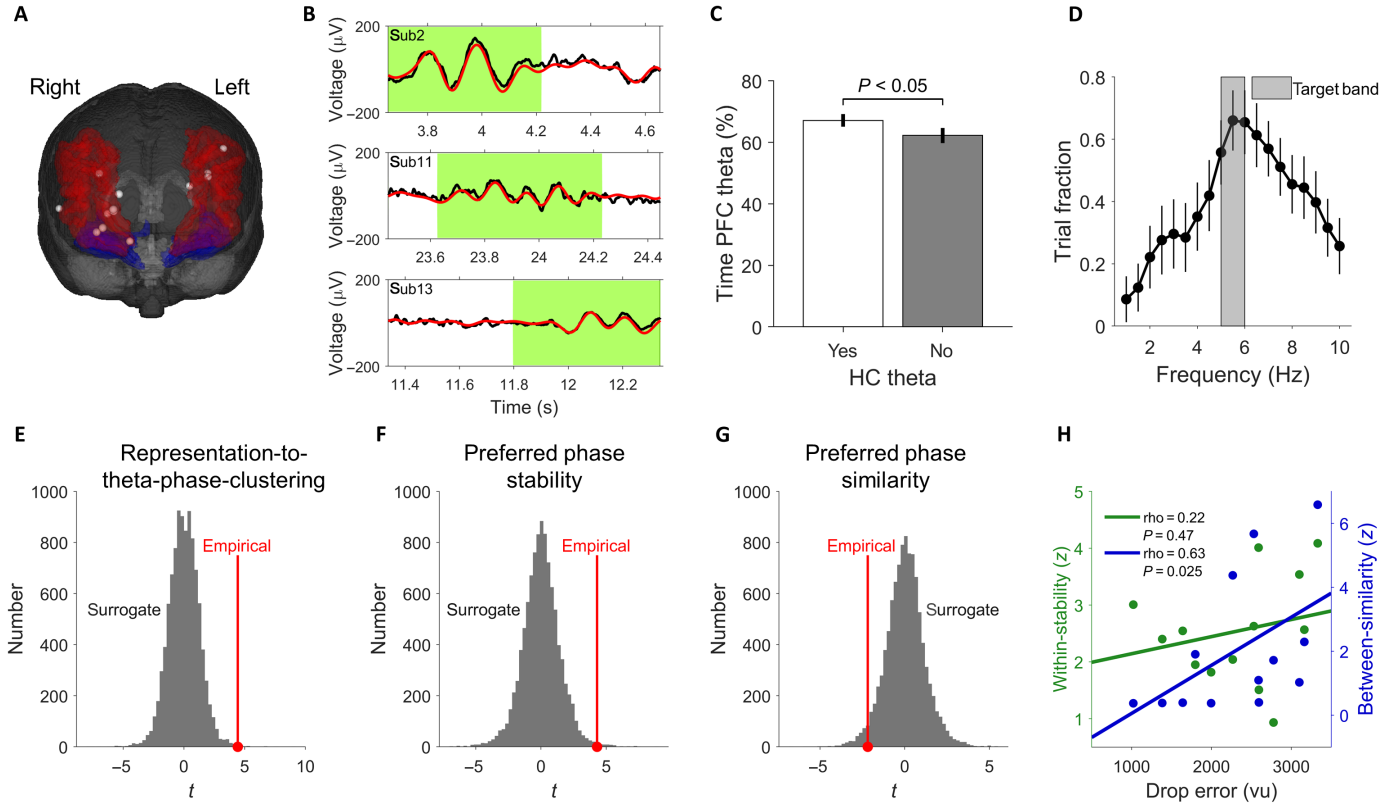

Fig. 6. Dynamic reactivation of cue representations at distinct prefrontal theta phases during goal-directed navigation. (**A**) Prefrontal electrode channels. Each white dot, one channel from a separate patient ($n = 13$). Blue, lateral orbitofrontal cortex; red, rostral middle frontal gyrus. (**B**) Exemplary time periods with theta oscillations during goal-directed navigation from different patients. Black, raw signal; red, low-frequency component of the raw signal (passband, 1 to 10 Hz); green shading, time periods with theta oscillations as detected by MODAL. (**C**) Prefrontal ("PFC") theta oscillations occurred in temporal proximity to hippocampal theta oscillations as shown by an increased percentage of time with prefrontal theta oscillations when hippocampal theta oscillations were present as compared to when they were not present. (**D**) Prefrontal theta oscillations preferentially occurred at a frequency of 5 to 6 Hz. Dots represent mean, and vertical lines represent SEM. (**E**) We assessed the significance of representation-to-theta-phase-clustering by comparing against a surrogate distribution that was obtained by circularly shifting the sliding RSA values against the concurrent prefrontal theta phases. Red dot and red line, empirical *t* statistic. (**F**) Stability of preferred theta phases across trials (within-stability). (**G**) Distinctiveness of preferred theta phases (between-similarity). (**H**) Lower between-similarity values are associated with better spatial memory performance.

In rodents, they appear as trains of rhythmical waves at around 4 to 10 Hz during movement in spatial environments. Both theta power and theta frequency are positively correlated with running speed (*14*), this relationship being diminished in virtual reality (*34*). In contrast to theta oscillations in rodents, human hippocampal theta oscillations typically occur in bursts lasting only a few cycles (*11*). Furthermore, they seem to occur at lower frequencies than in rodents (*13*), although this discrepancy might be attributable to virtual versus real-world navigation (*14*). The power of human hippocampal theta oscillations increases particularly during movement-onset periods and reflects the duration of the upcoming path, indicating their role in movement planning and execution (*11*, *35*). Theta oscillations do not only seem to be related to specific movement characteristics but may also reflect or even encode more complex features of the spatial environment. For example, theta oscillations in the human subiculum are increased during encoding of locations near boundaries (*36*), and theta oscillations in human entorhinal cortex exhibit characteristics of grid cells (i.e., sixfold rotational symmetry) (*37*, *38*). Beyond spatial navigation, theta oscillations play an important role during mnemonic processing (*39*). For example, 3-Hz "slow-theta" exhibits higher power and increased phase-amplitude coupling with gamma band activity during successful memory encoding (*40*).

The current study focused on hippocampal theta phases rather than on power and examined their relationship to large-scale neural representations of different mental contents. This framework is motivated by recent advances in identifying phase-dependent representations as a component of the human neural code (*21*, *30*). With this mechanism, a separation of otherwise interfering mental contents may be achieved, an idea that prompts discussion of the relationship between our findings and previous working memory studies.

As pointed out in the Introduction, goal-directed navigation shares commonalities with multi-item short-term memory tasks, in which different mental contents have to be separately maintained over a short period of time. A similar neural mechanism of theta-coupled reactivation of stimulus-specific neural representations (*4*) may underlie both cognitive functions. The allocation of neural representations to different subparts of a theta cycle has previously been suggested theoretically (*3*, *5*), and new evidence supports this theory, empirically showing that letter-selective broadband gamma activity sequentially occurs at different phases of the theta/alpha cycle (*6*). A similar finding was revealed during episodic sequence memory formation, showing that items at different sequence positions exhibit greater gamma power along distinct phases of a theta oscillation (*41*). Our study extends the coupling of different neural representations to

distinct theta phases to the complex behavior of goal-directed navigation and suggests that competing goals of a given context can be organized and separated within a theta cycle. Several previous short-term memory studies revealed a link between theta/gamma frequency ratio during phase-amplitude coupling and the individual short-term memory capacity (*42*). Similarly, future studies could examine the capacity of the hippocampal theta cycle during goal-directed navigation, i.e., how many different goals within a given context can be coded onto the theta cycle.

A difference of our study to these previous working memory studies is the theta frequency at which we observed behaviorally relevant locking of neural representations: Whereas working memory studies focused on frequencies in the theta/alpha frequency range (*6*) and 6 Hz in the classical theta frequency range (*4*), our study targeted lower theta frequencies at 3 to 4 Hz. Our approach was driven by the prevailing theta frequency during virtual navigation in our data and similar theta frequencies during virtual navigation in previous studies (*13*). Phase coding of neural representations to slow oscillations may thus be task specific, in turn explaining the difference between working memory capacity and the cognitive capacity of memorizing mental contents from larger temporal windows. An additional difference of our study to previous working memory studies is the specific way of extracting stimulus-specific neural representations. Whereas these previous studies obtained stimulus-specific neural representations from patterns of gamma power—e.g., at letter-selective cortical sites (*6*)—we detected cue-specific neural representations from large-scale electrophysiological time series data dominated by activity below 30 Hz. This analysis was motivated by recent advances in decoding stimuli from time series data (*17*) and the fact that time series data retain the rich information content of the original signal (*21*). Nevertheless, detailed analyses showed that neural cue representations could also be observed from gamma power patterns in our study and that their strength was correlated with the strength of neural cue representations based on the time domain data.

As revealed by our analysis relating the similarity of cue-specific neural representations to the subjective Euclidean distance between cue-associated goal locations, we find subtle evidence for spatial information content in these large-scale electrophysiological representations. Moreover, when we assessed the relevance of specific electrode channels to the overall representations via a jackknife procedure, lateral orbitofrontal cortex and rostral middle frontal gyrus particularly contributed to the representations. In combination, these two results parallel rodent and monkey studies, in which goals were represented in prefrontal cortex. For example, it has been shown that the rodent orbitofrontal cortex not only encodes information relating to abstract incentive value or general motivational significance but also codes for spatial goals defined as the concrete locations to which goal-directed navigation is directed (*9*). In a different study, prefrontal spiking activity patterns were identified to represent behavioral goals during spatial navigation strategies (*10*). Our results are in line with these findings and underline the importance of prefrontal cortices in goal-directed navigation. Notably, these regions—and particularly the lateral orbitofrontal cortex—were also found to be implicated in coding for stimulus-reward representations in previous studies (*43*). For example, lesions to the lateral orbitofrontal cortex lead to impairments in reward-credit assignment in macaques (*44*). More specifically, the lateral orbitofrontal cortex may be especially involved with assigning specific stimuli (or choices) to distinct types of reward, as demonstrated via functional magnetic resonance imaging (fMRI) in humans (*45*). Hence, the neural cue representations identified in our study may at least partially contain information about the reward pattern associated with each cue (although we only found a statistical trend in favor of this assumption in our data), in addition to information about the subjective distance between associated goal locations. Furthermore, it should be noted that the design of our task (being an associative cue-location memory task in which a given cue was always paired with the same goal location) did not allow us to disentangle neural representations that purely reflected spatial information of the associated goal locations from representations that purely reflected visual object properties of the cues. Future studies should thus further clarify the role of prefrontal regions during goal-directed navigation in humans.

In summary, our study identified large-scale electrophysiological representations of different objects that cued for associated goal locations. When tracking their reactivation during periods of goal-directed navigation, we found that representations of different cues locked to different hippocampal theta phases. Crucially, locking to more distinct theta phases was associated with better spatial memory performance. Our results therefore suggest that the hippocampal theta cycle provides a means to coherently represent different goals of a given context while preventing interference between them, thus shedding new light on the functional significance of theta oscillations (*46*). More generally, our results identify hippocampal theta phase coding as a neural mechanism underlying goal-directed navigation and open up new explanations for spatial disorientation as a symptom in neurological and psychiatric diseases.

MATERIALS AND METHODS

Experimental design

The objective of the current study was to evaluate the hypothesis that theta-coupled reactivation is a mechanism underlying goal-directed navigation in humans. To this end, we recorded iEEG from $N = 22$ epilepsy patients performing an object-location memory task navigating freely in a virtual environment. Cue-specific neural representations were extracted from the acquired brain-wide iEEG activity and, in a next step, related to the hippocampal theta rhythm that could be directly observed via hippocampal depth electrodes in a subset of our patients ($n = 16$).

Patients

Patients with medically intractable epilepsy ($N = 22$) participated in the current study (10 females; mean age ± SD, 29 ± 10 years; table S1). Patients underwent a surgical procedure in which electrodes were implanted subdurally on the cortical surface and/or stereotactically deep within the brain parenchyma. Electrode placements were determined solely on the basis of clinical considerations so as to best localize epileptogenic regions to guide respective treatment. Each patient had a unique implantation scheme with a unique anatomical distribution of electrode channels in various brain regions.

Recordings

iEEG recordings were performed at the Department of Epileptology, University of Freiburg, Freiburg im Breisgau, Germany; at the Epilepsy Centre Bethel, Bielefeld, Germany; at the Yuquan Hospital, Tsinghua University, Beijing, China; and at the First Affiliated Hospital

of PLA General Hospital, Beijing, China. Our research protocol was approved by the appropriate institutional review boards at each of the four hospital sites. Written informed consent was obtained from all patients. During recordings, all patients had normal or corrected-to-normal vision.

At the recording site in Freiburg, iEEG data were acquired using a Compumedics system (Compumedics, Abbotsford, Victoria, Australia) at a sampling rate of 2000 Hz. At the recording site in Bielefeld, iEEG data were acquired using a Nihon-Kohden system at a sampling rate of 1000 or 2000 Hz. At the recording sites in Beijing, iEEG data were acquired using a Nihon-Kohden system (Yuquan Hospital) and a Blackrock NeuroPort system (First Affiliated Hospital of PLA General Hospital) at a sampling rate of 2000 Hz. Electrodes were provided by Ad-Tech (Ad-Tech, Racine, WI, USA) at the recording sites in Freiburg and Bielefeld and by HKHS Beijing Health (HKHS Beijing Health Co., Ltd., Beijing, China) at the recording sites in Beijing. Signals were referenced to Cz (Freiburg), to linked mastoids (Bielefeld), or to one electrode contact located in white matter (Beijing). Regarding the latter, candidate reference electrode contacts located in white matter were chosen by visual inspection of the post-implantation computed tomography (CT) images coregistered onto the preimplantation MR images. Then, iEEG traces from each candidate reference electrode were visually inspected, and contacts with little or no apparent EEG activity were chosen as the reference for all subsequent recordings [see (*11*) for a similar procedure]. In total, signals from 2417 electrode channels distributed across varied brain regions were recorded.

Paradigm

The paradigm was adapted from a previous study (*16*). While being under continuous video EEG monitoring for diagnostic purposes, patients performed an object-location memory task navigating freely in a circular virtual environment. The environment comprised a grassy plane (diameter of 9500 vu) bounded by a cylindrical cliff. Two mountains, a sun, and several clouds provided patients with distal orientation cues rendered at infinity (fig. S1). No intramaze landmark was shown. Patients completed the task on a laptop using the arrow keys for moving forward and turning left and right and the spacebar or backward key to indicate their response. Patients were asked to complete up to 160 trials but were instructed to pause or quit the task whenever they wanted. At the very beginning, patients collected eight everyday objects (randomly drawn from a total number of 12 potential objects) from different locations in the arena ("initial learning phase"). Objects appeared one after the other. This time period (variable duration of approximately 2 min, as the whole task was self-paced) was excluded from all analyses. Afterward, patients completed variable numbers of trials, depending on compliance. Each trial consisted of four different phases (Fig. 1A). First, one of the eight objects was presented for 2 s (cue presentation). Afterward, patients were asked to navigate to the associated goal location within the virtual environment (retrieval). During the retrieval period, the cue image was not present anymore. There was no delay period between cue presentation and the retrieval period. After patients had indicated their response via a button press at the assumed goal location, they received feedback depending on response accuracy (feedback; fixed duration of 1.5 s). Response accuracy was measured as the distance between the assumed goal location and the correct goal location (drop error). Last, the object was presented in the correct location, and patients had to collect the object to further improve their associative memory between the object and its goal location (re-encoding). After each trial, a fixation crosshair was shown for a variable duration of 3 to 5 s (uniformly distributed). Across trials, patients had to retrieve the cue objects in random order, preventing them from using a sequential learning strategy. Furthermore, starting locations were identical with ending locations from preceding trials and thus varied from trial to trial, preventing patients from using a response-based navigation strategy. Chance performance of drop errors was determined by randomly assigning response locations to correct goal locations 50,000 times per patient and averaging across trials, surrogate repetitions, and patients afterward to obtain one overall chance level value. Experimental events were written to a log file (temporal resolution of 20 ms). Speed was calculated as $v = d/t$, where d is the distance between consecutive locations within the virtual environment and t is the duration between corresponding time stamps. Triggers were either detected using a phototransistor attached to the screen marking onsets and offsets of the cue presentation phase or using an independent custom MATLAB (2017b, The MathWorks, Massachusetts) program that sent triggers both to the paradigm and to the iEEG recording software with randomly jittered intervals between 0.5 and 5 s. All of our analyses focused on the cue presentation and the retrieval period.

Identification of channel locations

For patients from Freiburg and Bielefeld, for whom one preimplantation and one post-implantation MRI were available, electrode localization was performed using FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL) and PyLocator (http://pylocator.thorstenkranz.de/). The post-implantation MR image was coregistered with the preimplantation MR image. Next, the preimplantation MR image was skull-stripped and normalized to MNI space, and the same normalization matrix was applied to the post-implantation MR image. Normalized post-implantation images were visually inspected using PyLocator, and channel locations were manually identified. For patients from Beijing, for whom one preimplantation MR image and one post-implantation CT scan were available, electrode localization in MNI space was performed using a custom toolbox (*47*).

To assign electrode channel MNI coordinates to brain regions, we first subjected the average structural template image from FSL to the FreeSurfer pipeline (https://surfer.nmr.mgh.harvard.edu/), providing one cortical/subcortical label for each MNI coordinate. Next, we went through each electrode channel and assigned the closest cortical/subcortical label. This procedure then allowed us to select electrode channels from specific regions of interest and to assign the jackknife resampling results (see below) to different cortical/subcortical regions.

Preprocessing

Channels containing obvious artifacts were excluded based on visual inspection, leading to a final number of 2330 usable recording channels (dropout of 3.6%). With obvious artifacts, we refer to implausibly high absolute voltages or implausibly high-voltage gradients that are present in electrode channels presumably located outside the brain. Because trial numbers per patient were limited to 40 to 160 (5 to 20 per individual cue), we dealt with potential epileptic activity using the following two strategies. First, regarding RSA, we did not exclude any trials, because artifact correction is less critical when applying multivariate analyses (*17*). To attenuate the impact of potential outliers, we applied nonparametric Spearman correlations throughout

(both for measuring neural similarity and for second-level statistics). Second, we reperformed the hippocampal representation-to-theta-phase-clustering analysis using a restricted subset of $n = 12$ patients (excluding 4 of the 16 patients because of potentially higher levels of epileptic activity in their hippocampal channels, as revealed by visual inspection of the hippocampal ERPs) to underscore that our results were not critically influenced by epileptic activity.

Preprocessing was performed on the entire original data using FieldTrip (*22*) (www.ru.nl/neuroimaging/fieldtrip) and included filtering (high-pass filtering with a frequency of 0.1 Hz; low-pass filtering with a frequency of 200 Hz; band-stop filtering with frequencies of 50, 100, and 150 Hz) and resampling to 1000 Hz.

Time-resolved sRSA

Time-specific decoding is becoming increasingly popular for magnetencephalography (MEG) and EEG analyses (*17*). Here, we used tr-sRSA as the decoding approach. The principal assumption that underlies this RSA approach is that identical stimuli show stronger neural similarity (as measured by Spearman correlations) than different stimuli. RSA (*24*) is well suited in cases when only relatively few samples per category are available (because no training data are needed). In our case, only two samples per cue were available because we increased the signal-to-noise ratio (SNR) before RSA by time point–specific averaging across repetitions of the same cue after randomly distributing trials onto two different data halves.

Preprocessed data were low-pass–filtered at 30 Hz and epoched using a time window of −2 to +2 s with respect to cue presentation onset. ICA (using FieldTrip's runica implementation with default settings) was then run on the epoched data to enhance local information, similar or superior to bipolar referencing (*20*). Roughly speaking, ICA demixes the channel data into information sources. Because each patient had a unique implantation scheme and was thus implanted with a unique number n_i of channels, the number of independent components c_i also varied between patients. Next, we baseline-corrected the epoched components by subtracting the mean of the period between −0.2 and 0 s relative to cue presentation onset. Trials were then randomly distributed onto two data halves. Within each data half, we calculated one NV per cue by averaging across trials of the same cue, separately for each time point within the epoch. Averaging across trials was done to increase the SNR, as suggested previously (*17*). Hence, separately for each time point during the epoch, we obtained one component × 1 NV for each cue. Afterward, neural similarity was assessed by calculating the Fisher-*z*–transformed Spearman correlation coefficient between all combinations of NV_i and NV_j, where i is the cue index of the first data half and j is the cue index of the second data half. Separately for each time point during the epoch, this resulted in an 8 × 8 confusion matrix of neural similarities between identical (on-diagonal) and different (off-diagonal) cues. We chose nonparametric Spearman correlations over parametric Pearson correlations to attenuate the potential influence of outliers consistent with previous RSA studies [e.g., (*19*, *48*)]. To decrease a potential influence of the random assignment to the two different data halves, we repeated the procedure of calculating the confusion matrices 10 times and averaged over repetitions to obtain a final confusion matrix. The eight on-diagonal values of the final confusion matrix containing the neural similarity values of identical cues were averaged, providing a time point–specific measure of neural similarity for identical cues ($S_{identical}$). From each row of the confusion matrix, we then randomly selected one neural similarity value representing the similarity between different cues. Again, these eight values were averaged, now providing a time point–specific measure of neural similarity between different cues ($S_{different}$). A similar number of on- and off-diagonal values were used to avoid effects of regression toward the mean. Hence, for each patient, we obtained one $S_{identical}$ and one $S_{different}$ time course between −2 and +2 s relative to cue presentation onset.

Statistical differences between the time courses of $S_{identical}$ (one time course per patient) and $S_{different}$ (one time course per patient) were assessed using cluster-based permutation testing (*22*). In detail, we calculated a paired t test across patients between $S_{identical}$ and $S_{different}$, separately for each time point. Contiguous clusters of significant t values at an uncorrected P value of <0.05 were identified and summed up, resulting in one sum t value per cluster. Only the highest sum t value was kept ($t_{empirical}$). To create surrogate sum t values ($t_{surrogate}$), we randomly switched $S_{identical}$ and $S_{different}$ per patient and applied the identical statistical procedure, as described above. One thousand $t_{surrogate}$ values were created and compared to $t_{empirical}$ by assessing the percentile of $t_{empirical}$ within $t_{surrogate}$. The 95th percentile was defined as statistically significant at an alpha level of $P < 0.05$. The significant time period defined a tROI for subsequent analyses (see below). This tROI was determined across patients (revealing during which time period identical cues generally elicited higher neural similarity values than different cues), but prototypal neural cue representations were specific to each patient, because they were determined separately for each patient by averaging the voltage values of the c_i independent components across trials of the same cue and across the tROI, resulting in a unique $1 \times c_i$ voltage vector per prototypal neural cue representation per patient.

To demonstrate the specificity of neural cue representations, we calculated neural cue representations separately for both data halves (i.e., we averaged the voltage values of the c_i independent components across trials of the same cue—after distributing trials onto the two data halves—and across the tROI). Neural cue representations were ordered as a function of the associated subjective goal location (see the "Information content of large-scale electrophysiological cue representations" section and fig. S2). For each patient, we then calculated the neural similarity (Spearman correlation) between all pairs of neural cue representations from both data halves, resulting in an 8 × 8 neural similarity matrix, where on-diagonal values contain neural similarities of identical cues from both data halves ($SIM_{identical}$) and off-diagonal values contain neural similarities of different cues from both data halves ($SIM_{different}$). For graphical depiction, we averaged neural similarity values across subjects [following (*23*)]. For statistical evaluation, we calculated the percentage of $SIM_{identical}$ being higher than $SIM_{different}$, separately for each patient and each cue. Using one-sample t tests across subjects, we evaluated whether percentage values were above chance level (50%) for each cue.

Time-resolved spatial multivariate pattern analysis

Parallel to our tr-sRSA procedure, preprocessed data were low-pass–filtered (30 Hz) and epoched using a time window of −2 to +2 s with respect to cue presentation onset. Epoched data were subjected to ICA and baseline-corrected. For each time point, we then constructed a component × 1 NV across ICA components for each trial, giving a $t \times f$ matrix, where t is the number of trials and f is the number of features (i.e., ICA components), separately for each time point during cue presentation. A 10-fold cross-validation regime was then used to decode the eight different cues based on their associated

NVs. In detail, during each of the cross-validation runs, 90% of the trials were used to train a 10–nearest neighbor classifier (MATLAB's fitcknn; distance metric, Spearman's rho), whose classification model was then used to predict the cue labels in the remaining 10% of trials based on the associated NVs. This procedure resulted in a time course of empirical classifier accuracy values ($CA_{empirical}$). In addition, a time course of surrogate classifier values ($CA_{surrogate}$) was created by circularly shifting cue labels across trials by a random integer. Statistical differences between the time courses of $CA_{empirical}$ (one time course per patient) and $CA_{surrogate}$ (one time course per patient) were assessed using cluster-based permutation testing.

tr-sRSA based on gamma power

We epoched the data from 2 s before cue onset until 2 s after cue offset and converted the channel-wise data into independent components (a larger time window was chosen to exclude edge artifacts, resulting from time-frequency decomposition). Next, we extracted gamma power via Morlet wavelets (seven cycles) from 30 to 90 Hz, in steps of 4 Hz. Trial epochs were then reduced to 0.1 s before cue onset until 0.1 s after cue offset. Gamma power values were *z*-scored (i.e., normalized by first subtracting the average value and then dividing by the standard deviation across all trials), separately for each frequency and each independent component. Next, tr-sRSA was performed as described above, resulting in a time point– and frequency-specific measure of neural similarity for identical cues ($S_{identical}$) and a time point– and frequency-specific measure of neural similarity between different cues ($S_{different}$). Hence, for each patient, we obtained one $S_{identical}$ and one $S_{different}$ frequency by time matrix between −0.1 and +2.1 s relative to cue presentation onset. Statistical differences between $S_{identical}$ and $S_{different}$ were assessed using cluster-based permutation testing.

Information content of large-scale electrophysiological cue representations

To understand the information content of the large-scale electrophysiological cue representations in greater detail, we tested the hypothesis that the pairwise similarity between cue representations is related to the pairwise similarity of objective/subjective goal locations (objective goal locations are the true goal locations; subjective goal locations are the average response location of each cue). To this end, we first calculated the neural similarity matrix between all possible pairs of cue representations (for an example, see fig. S4A) and the goal-location similarity matrix between all possible pairs of objective/subjective goal locations (for an example, see fig. S4, B and C). Goal-location similarity is estimated via a linear transformation of the Euclidean distance between a given pair of cue-specific goal locations: goal-location similarity $(cues_{ij}) = 1 - D_{ij}/\max(D)$, where D_{ij} is the distance between the goal locations of cue *i* and cue *j*. Next, we computed the higher-order similarity by correlating the neural similarity values with the goal-location similarity values (Spearman correlation). This resulted in one Spearman's rho value per subject evaluating whether pairs of cues with similar neural representations exhibited more (or less) similar objective/subjective goal locations (for an example, see fig. S4D). Subject-specific higher-order rho values were then Fisher-*z*–transformed and fed into a one-sample *t* test across subjects to test for a consistent relationship between neural similarity matrices and goal-location similarity matrices across subjects.

An analogous procedure was chosen to test for other factors potentially contributing additional information content to the cue representations or potentially constituting confounding variables: cue-specific (i) movement direction patterns, (ii) speed patterns, (iii) acceleration patterns, (iv) visual properties of the cue objects, and (v) reward patterns. For the cue-specific movement direction patterns, we calculated the pairwise similarities via Kuiper's tests. For the cue-specific speed patterns, acceleration patterns, and reward patterns, we calculated the pairwise similarities via two-sample Kolmogorov-Smirnov tests. Test statistics were linearly transformed into similarity values via $k' = 1 - k/\max(k)$. Higher-order similarity between neural similarity values and similarity values k' was then computed as outlined above. Regarding visual properties of the cue objects, we evaluated each cue object at different levels of visual complexity via the DNN "AlexNet" (*49*) with eight layers (five convolutional layers and three fully connected layers). The DNN was implemented using the Caffe framework in Python and pretrained on the ImageNet dataset. For example, the first layer represents basic visual properties of the cue objects such as colors and edges. Next, we calculated the pairwise visual similarity of the cue objects at each DNN layer (Spearman correlations) to finally compute higher-order similarities with the neural similarity matrices as outlined above.

To examine whether cue representations underwent general changes due to learning, we tested whether cue representations were more distinct from each other at the end (last trial chunk) as compared to the beginning of the task (first trial chunk). With "trial chunk," we refer to the collection of eight consecutive trials, in which each cue has to be retrieved once. Hence, for both the first and the last trial chunk, we extracted neural cue representations (i.e., we averaged NVs within the tROI from Fig. 2C) and correlated them with each other (separately for the two trial chunks), resulting in two neural similarity matrices. In these similarity matrices, the lower triangle contains the similarity values between different cue representations. We averaged the lower triangle of each similarity matrix, giving one average neural similarity value for both trial chunks and for each patient. We tested across patients whether neural similarity values were lower in the last trial chunk than in the first trial chunk.

Contribution of different brain regions to the neural cue representations

To assess the contribution of single channels to the RSA results, we calculated two "contribution" scores C_{on} and C_{off} for each channel *c*. Whereas C_{on} quantifies how much a given channel increases the neural similarity of identical cues from both data halves ("on" standing for "on-diagonal values," reflecting the neural similarity of identical cues), C_{off} quantifies how much a given channel increases the neural similarity of different cues from both data halves ("off" standing for "off-diagonal values," reflecting the neural similarity of different cues). C_{on} and C_{off} values were obtained by performing a jackknife resampling procedure during which the identical RSA procedure (including the ICA transformation from iEEG channel space into ICA component space) as explained above was completed *n* times, where *n* refers to the number of channels. During each round, one channel *c* was left out.

To compute C_{on} for a given channel *c*, the AUC (AUC_c) of $S_{identical}$ (i.e., the area between $S_{identical}$ and the *x* axis) during the tROI was calculated. This AUC_c was then subtracted from the original AUC_{all} that was determined during the tROI using all channels. A positive difference between AUC_{all} and AUC_c (i.e., $AUC_{all} - AUC_c > 0$) indicates that channel *c* contributed positively to the original RSA

result (with C_{on} = AUC_{all} − AUC_c). For each brain region (as given by FreeSurfer's cortical parcellation output), we determined the percentage of channels with a positive C_{on} value (termed "relative engagement"; Fig. 2E). Note that during this analysis, contribution scores C_{on} were concatenated across patients, resulting in a statistical evaluation across channels. To assess the significance of region-specific relative engagement, we shuffled the contribution scores C_{on} relative to the region labels 2000 times, each time obtaining one surrogate relative engagement value per brain region. Empirical relative engagement values were then compared against the surrogate relative engagement values. Brain regions whose empirical relative engagement values exceeded the 95th percentile of surrogate relative engagement values were considered significant after correcting for 29 different brain regions (Bonferroni correction).

To compute C_{off} for a given channel c, the AUC (AUC_c) of $S_{different}$ (i.e., the area between $S_{different}$ and the x axis) during the tROI was calculated. This AUC_c was then subtracted from the original AUC_{all} that was determined during the tROI using all channels. A positive difference between AUC_{all} and AUC_c (i.e., $AUC_{all} - AUC_c > 0$) indicates that channel c contributed positively to the neural similarity of different cues from both data halves (thus reducing the strength of the tr-sRSA result). For each brain region, we determined the percentage of channels with a positive C_{off} value, where a low percentage value represents a beneficial contribution to discriminate between the neural representations of different cues. Statistical evaluation proceeded as described for C_{on}, with the difference that brain regions whose empirical proportion fell below the fifth percentile of surrogate proportions (because brain regions reducing the neural similarity of different cues were of interest) were considered significant after correcting for 29 different brain regions (Bonferroni correction). Because no brain region fulfilled this strict statistical criterion, we plot the result at an uncorrected statistical threshold of $P < 0.05$ for illustrative purposes only (fig. S6).

To statistically evaluate the contribution of different brain regions to the significant difference between $S_{identical}$ and $S_{different}$ in the tr-sRSA across patients, we first calculated the contribution score C_{on} and the contribution score C_{off} for each channel (see above). Both C_{on} and C_{off} are relevant for establishing a significant difference in our tr-sRSA: A strong positive difference between C_{on} and C_{off} (i.e., $C_{on} - C_{off} \gg 0$) means that a given channel strongly contributes to a significant difference between $S_{identical}$ and $S_{different}$. Next, we tagged each voxel within a radius of 12.5 mm around a given channel MNI coordinate with the channel-specific ($C_{on} - C_{off}$) value. For each patient, we averaged across values if multiple channels contributed to a given MNI coordinate. This gave one three-dimensional ($C_{on} - C_{off}$) map in MNI space per patient. Across patients, we then performed a t test against zero, separately for each voxel (fig. S7). Statistical evaluation of the t map was performed via cluster-based permutation testing (*22*). Contiguous three-dimensional clusters of significant t values at $P_{unc.} < 0.05$ were identified and summed up, resulting in a cluster t value for each cluster. Empirical cluster t values were compared against 1000 surrogate cluster t values obtained by randomly flipping C_{on} and C_{off} volumes (separately for each patient) before identifying contiguous clusters of significant t values at $P_{unc.} < 0.05$. In each surrogate round, the cluster with the highest cluster t value was kept. Empirical cluster t values were then ranked within the 1000 surrogate cluster t values (a rank of >950 designates an empirical cluster t value as significant). The identified cluster in prefrontal cortex served as an anatomical criterion for the selection of prefrontal channels in subsequent analyses.

Hippocampal ERP analysis

Preprocessed iEEG data were low-pass–filtered at 30 Hz. Data epochs between −0.5 and 2 s with respect to cue presentation onset were extracted for each trial. Trial-wise baseline correction was performed by subtracting the mean during a prestimulus interval from −0.2 to 0 s with respect to cue presentation onset. For all patients who had at least one hippocampal channel (n = 16), grand-average ERPs across all trials from all hippocampal channels were visually inspected, and the channel with the clearest ERP was selected. More detailed, electrode channels were selected based on several criteria that were applied in consecutive steps. First, we selected channels that were located in the anterior hippocampus. Second, we only included channels that showed a negative main deflection. Third, we identified the channel with the ERP showing the highest SNR, which was computed as follows: SNR = max(abs(V_{cue}))/std($V_{baseline}$), where V_{cue} are the voltage values during cue presentation and $V_{baseline}$ are the voltage values during the baseline period. Fourth, when two channels had a similar SNR, their grand-average ERPs were visually inspected, and the one with the higher signal quality (i.e., a lower amount of super-imposed high-frequency artifacts) was selected. Notably, these criteria were all applied before the consecutive analysis steps, avoiding any bias in the selection of electrodes. Because iEEG data from epilepsy patients can be distorted by epileptic activity, we also performed the analysis of the dynamic reactivation of large-scale electrophysiological cue representations at specific hippocampal theta phases using a restricted subset of patients (n = 12) who showed the clearest ERPs during cue presentation in comparison to the other patients as a control analysis.

Analysis of representation-to-theta-phase-clustering during retrieval periods

For each cue, we extracted a prototypal neural representation by averaging the component-wise iEEG signal within the tROI and then across trials, resulting in one component × 1 prototypal neural representation per cue, separately for each patient. The number of components varied between patients due to the fact that each patient had a unique implantation scheme. Hence, cue-specific prototypal neural representations also varied between patients with respect to the anatomical distribution of their neural sources. The number of components is identical with the number of electrode channels, and no components were excluded before the analysis of representation-to-theta-phase-clustering. The entire preprocessed data were then subjected to the previously calculated ICA unmixing matrix converting the original channel-wise data into component-wise data. Component-wise data were epoched into trials and baseline-corrected. We then slid the prototypal cue representation whose goal location had to be retrieved during a given trial across all NVs during the retrieval period of this trial, resulting in a time series of sliding RSA values for each trial presumably representing dynamic reactivation of the prototypal cue representation (Fig. 4, A and B). This dynamic reactivation is specific to each patient and each trial and depicts how much the patient's current brain state resembles the prototypal neural cue representation formed during the tROI of the cue presentation period whose associated goal location shall be retrieved in this trial.

Hippocampal theta phases were extracted from the preprocessed iEEG data of the hippocampal channel that was selected during the ERP analysis (see above). To define the exact frequency from which to extract theta phase information, we opted for a data-driven method using MODAL (*30*) that dynamically detects narrow-band

oscillations exceeding the background 1/*f* spectrum. Only bands with a lower border of ≥1 Hz and an upper border of ≤10 Hz based on previous results (*11*) were kept, and a summary plot of the extracted theta bands that showed a peak at 3.5 Hz, which occurred in 61.3% of the extracted bands [similar to (*13*, *30*)], was created (Fig. 3D). For subsequent analyses, we then focused on the oscillatory phase around this predominant frequency of 3.5 Hz. In greater detail, during all epochs that were identified by MODAL to contain oscillatory activity in the theta range, we bandpass-filtered the data between 3 and 4 Hz and extracted the instantaneous phase via a Hilbert transformation. This target frequency range of 3 to 4 Hz is part of "slow" theta oscillations that have been shown to exhibit increased power during successful memory encoding (*40*) and seem to be particularly prevalent during (virtual) spatial navigation in humans (*11*, *13*, *14*).

For each patient, we then assessed the cue representation-specific clustering of preferred phases across trials. To this end, we related the cue representation-specific time series of dynamically changing RSA values to the phases of the hippocampal theta oscillation via the Moore-Rayleigh test (which is a nonparametric extension of the Rayleigh test weighting the input phases by a ranked factor, i.e., the sliding RSA values in our case) (*31*). This results in one preferred theta phase per trial. We tested the significance of the representation-to-theta-phase-clustering by comparing the Moore-Rayleigh's r^* values obtained with the original data with surrogate Moore-Rayleigh's r^* values obtained by circularly shifting the RSA values in relation to the concurrent hippocampal theta phases. In detail, we obtained one empirical Moore-Rayleigh's r^* value and 1000 surrogate Moore-Rayleigh's r^* values for each trial of each patient. Empirical r^* values were then averaged across trials, resulting in one mean empirical r^* value per patient. With respect to surrogate r^* values, one surrogate r^* value was randomly selected per trial (we did not average across all 1000 surrogate values in order to avoid regression toward the mean). These randomly selected r^* values were then averaged across trials, resulting in one mean surrogate r^* value per patient. Using a paired *t* test, we evaluated whether mean empirical r^* values were higher than mean surrogate r^* values, resulting in an empirical *t* statistic $t_{empirical}$. To assess the significance of this *t* value, we compared it to a surrogate distribution of *t* values that were created by randomly swapping mean empirical r^* values and mean surrogate r^* values 1000 times and recomputing the paired *t* test in each round. This procedure results in a distribution of surrogate *t* values $t_{surrogate}$, in which the rank of $t_{empirical}$ can be determined. The corresponding *P* value can be computed as $P = 1 - rank/1000$. Then, for each trial, we extracted the preferred phase for the cue whose goal location had to be retrieved during this trial. Consistency of preferred phases across trials (within-stability) was assessed separately for each of the eight cue representations using Rayleigh's tests, leading to one empirical *z* value per cue that indicates the circular clustering of preferred phases for the corresponding cue representation. Rayleigh's *z* values were averaged afterward, resulting in one $z_{within\text{-}stability}$ value per patient. The higher $z_{within\text{-}stability}$, the more consistent the preferred theta phases were across trials of the same cue. Statistical significance was assessed by comparing $z_{within\text{-}stability}$ with surrogate $z_{within\text{-}stability}$ values that were obtained by randomly shuffling which cue was shown during a given trial. In detail, we obtained one empirical $z_{within\text{-}stability}$ value per patient (correct assignment which cue had to be retrieved during a given trial) and 1000 surrogate $z_{within\text{-}stability}$ values per patient (random assignment which cue had to be retrieved during a given trial in each round). For each patient, one surrogate value was randomly selected (we did not average across all 1000 surrogate values to avoid regression toward the mean). Using a paired *t* test, we then evaluated whether empirical $z_{within\text{-}stability}$ values were higher than surrogate $z_{within\text{-}stability}$ values, resulting in an empirical *t* statistic $t_{empirical}$. To assess the significance of this *t* value, we compared it to a surrogate distribution of *t* values that were created by randomly swapping empirical and surrogate $z_{within\text{-}stability}$ values 1000 times and recomputing the paired *t* test in each round. This procedure results in a distribution of surrogate *t* values $t_{surrogate}$, in which the rank of $t_{empirical}$ can be determined. Again, the corresponding *P* value can be computed as $P = 1 - rank/1000$.

In addition to the cue-specific clustering across trials, we also assessed the clustering of preferred phases between different cue representations (between-similarity), asking whether different cue representations clustered at different theta phases. To this end, the eight mean preferred theta phases (one for each cue representation) were subjected to a Rayleigh test separately for each patient, resulting in $z_{between\text{-}similarity}$. The lower $z_{between\text{-}similarity}$, the more different the representation-specific preferred theta phases. Empirical $z_{between\text{-}similarity}$ values were compared with surrogate $z_{between\text{-}similarity}$ values obtained by randomly assigning cues to trials. Statistical evaluation proceeded exactly as described for the $z_{within\text{-}stability}$ values. Because we tested whether empirical $z_{between\text{-}similarity}$ values were significantly smaller than surrogate $z_{between\text{-}similarity}$ values, the final *P* value was calculated as $P = rank/1000$. To assess a behavioral relevance of representation-to-theta-phase-clustering, we calculated Spearman correlations between the mean drop error and, on the one hand, $z_{within\text{-}stability}$ and, on the other hand, $z_{between\text{-}similarity}$ across patients.

To present evidence that our results of representation-to-theta-phase-clustering were not due to trial-by-trial variations of the predominant theta frequencies, we examined whether instantaneous theta frequencies varied as a function of retrieved cue. Instantaneous theta frequencies were determined via the "frequency sliding" approach (*50*): We bandpass-filtered the raw signal from the hippocampal channel within 1 to 10 Hz and calculated instantaneous phases via Hilbert transformation. We then extracted instantaneous frequencies following Cohen (*50*) and averaged instantaneous frequencies within trials. For each patient, we then tested whether average theta frequencies varied as a function of retrieved cue using one-way ANOVAs across cues.

The analysis of representation-to-theta-phase-clustering during retrieval periods with respect to prefrontal cortex theta oscillations proceeded exactly as described above with the difference that prefrontal theta oscillations exhibited a higher peak frequency (5.5 Hz). Thus, during all epochs that were identified by MODAL to contain oscillatory activity in the theta range, we bandpass-filtered the data between 5 and 6 Hz and extracted the instantaneous phase via Hilbert transformation to associate the cue representation-specific time series of dynamically changing RSA values to the phases of the prefrontal theta oscillations.

Phase coupling between hippocampus and prefrontal cortex

Because previous work in rodents showed enhanced coupling between hippocampus and prefrontal cortex during task periods associated with peak mnemonic and decision-making load, we examined phase coupling between hippocampus and prefrontal cortex at the end of goal-directed navigation when decision-making load presumably was highest (last 1.5 s of the retrieval period). The time window of 1.5 s was chosen as a good balance between the

specificity of the ongoing behavior and the number of theta cycles contained in this period (more than two full cycles at the lowest frequency). Phase coupling was computed via the PLV across time, separately for each trial (*33*). Trial-wise PLVs were estimated for a range of frequencies (1.5 to 10 Hz, in steps of 0.5 Hz) after extracting hippocampal and prefrontal phase values via Hilbert transformation (including 1-Hz bandpass filtering using MATLAB's firls). For each patient, we computed the average PLV during good and during bad trials and compared the two conditions across patients via a paired *t* test. *P* values were Bonferroni-corrected for the number of frequencies. As a control analysis, we compared empirical PLVs against surrogate PLVs that were obtained by randomly assigning prefrontal phase time series to trials.

Statistical analysis

All analyses were performed in MATLAB 2017b using custom MATLAB scripts and toolboxes, as outlined above. Inference statistics across patients were performed using MATLAB or SPSS (version 24.0, IBM, NY). Types of statistical tests used are specified where the test statistics are reported. Statistical analyses were performed using a significance threshold of $P < 0.05$. All analyses were two-tailed, if not otherwise specified. If appropriate, Bonferroni correction for multiple comparisons was applied by multiplying the output *P* value of the test statistic by the number of tests performed. Error bars in figures are defined in the corresponding figure legends. We report degrees of freedom in *t* tests and the number of subjects in correlation analyses. All statistical analyses are described in detail in Results and Materials and Methods.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/7/eaav8192/DC1

Table S1. Patient information.
Table S2. MNI coordinates of hippocampal channels.
Fig. S1. Layout of the virtual environment and patient-wise goal locations.
Fig. S2. Stimulus specificity of neural cue representations.
Fig. S3. Identification of large-scale electrophysiological cue representations using tr-sRSA based on gamma power.
Fig. S4. Derivation of higher-order similarity.
Fig. S5. Neural cue representations rely on large-scale electrophysiological signals.
Fig. S6. Contribution of brain regions to the similarity of identical and different large-scale electrophysiological cue representations.
Fig. S7. Second-level statistics across patients depicting which brain regions simultaneously increased the neural similarity of identical cues and decreased the neural similarity of different cues.
Fig. S8. Phase coupling (3.5 Hz) between the hippocampus and prefrontal cortex during goal-directed navigation.

REFERENCES AND NOTES

1. F. Chersi, N. Burgess, The cognitive architecture of spatial navigation: Hippocampal and striatal contributions. *Neuron* **88**, 64–77 (2015).
2. M. D'Esposito, B. R. Postle, The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* **66**, 115–142 (2015).
3. J. E. Lisman, M. A. Idiart, Storage of 7 +/- 2 short-term memories in oscillatory subcycles. *Science* **267**, 1512–1515 (1995).
4. L. Fuentemilla, W. D. Penny, N. Cashdollar, N. Bunzeck, E. Düzel, Theta-coupled periodic replay in working memory. *Curr. Biol.* **20**, 606–612 (2010).
5. J. E. Lisman, O. Jensen, The θ-γ neural code. *Neuron* **77**, 1002–1016 (2013).
6. A. Bahramisharif, O. Jensen, J. Jacobs, J. Lisman, Serial representation of items during working memory maintenance at letter-selective cortical sites. *PLOS Biol.* **16**, e2003805 (2018).
7. W. E. Allen, I. V. Kauvar, M. Z. Chen, E. B. Richman, S. J. Yang, K. Chan, V. Gradinaru, B. E. Deverman, L. Luo, K. Deisseroth, Global representations of goal-directed behavior in distinct cell types of mouse neocortex. *Neuron* **94**, 891–907.e6 (2017).
8. H. T. Ito, S.-J. Zhang, M. P. Witter, E. I. Moser, M.-B. Moser, A prefrontal–thalamo–hippocampal circuit for goal-directed spatial navigation. *Nature* **522**, 50–55 (2015).
9. C. E. Feierstein, M. C. Quirk, N. Uchida, D. L. Sosulski, Z. F. Mainen, Representation of spatial goals in rat orbitofrontal cortex. *Neuron* **51**, 495–507 (2006).
10. I. Negrón-Oyarzo, N. Espinosa, M. Aguilar-Rivera, M. Fuenzalida, F. Aboitiz, P. Fuentealba, Coordinated prefrontal–hippocampal activity and navigation strategy-related prefrontal firing during spatial memory formation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7123–7128 (2018).
11. D. Bush, J. A. Bisby, C. M. Bird, S. Gollwitzer, R. Rodionov, B. Diehl, A. W. McEvoy, M. C. Walker, N. Burgess, Human hippocampal theta power indicates movement onset and distance travelled. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12297–12302 (2017).
12. V. D. Bohbot, M. S. Copara, J. Gotman, A. D. Ekstrom, Low-frequency theta oscillations in the human hippocampus during real-world and virtual navigation. *Nat. Commun.* **8**, 14415 (2017).
13. A. J. Watrous, D. J. Lee, A. Izadi, G. G. Gurkoff, K. Shahlaie, A. D. Ekstrom, A comparative study of human and rat hippocampal low-frequency oscillations during spatial navigation. *Hippocampus* **23**, 656–661 (2013).
14. Z. M. Aghajan, P. Schuette, T. A. Fields, M. E. Tran, S. M. Siddiqui, N. R. Hasulak, T. K. Tcheng, D. Eliashiv, E. A. Mankin, J. Stern, I. Fried, N. Suthana, Theta oscillations in the human medial temporal lobe during real-world ambulatory movement. *Curr. Biol.* **27**, 3743–3751.e3 (2017).
15. S. Fujisawa, G. Buzsáki, A 4-Hz oscillation adaptively synchronizes prefrontal, VTA, and hippocampal activities. *Neuron* **72**, 153–165 (2011).
16. L. Kunz, T. N. Schröder, H. Lee, C. Montag, B. Lachmann, R. Sariyska, M. Reuter, R. Stirnberg, T. Stöcker, P. C. Messing-Floeter, J. Fell, C. F. Doeller, N. Axmacher, Reduced grid-cell–like representations in adults at genetic risk for Alzheimer's disease. *Science* **350**, 430–433 (2015).
17. T. Grootswagers, S. G. Wardle, T. A. Carlson, Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J. Cogn. Neurosci.* **29**, 677–697 (2017).
18. Y. Lu, C. Wang, C. Chen, G. Xue, Spatiotemporal neural pattern similarity supports episodic memory. *Curr. Biol.* **25**, 780–785 (2015).
19. H. Zhang, J. Fell, B. P. Staresina, B. Weber, C. E. Elger, N. Axmacher, Gamma power reductions accompany stimulus-specific representations of dynamic events. *Curr. Biol.* **25**, 635–640 (2015).
20. S. Michelmann, M. S. Treder, B. Griffiths, C. Kerrén, F. Roux, M. Wimber, D. Rollings, V. Sawlani, R. Chelvarajah, S. Gollwitzer, G. Kreiselmeyer, H. Hamer, H. Bowman, B. Staresina, S. Hanslmayr, Data-driven re-referencing of intracranial EEG based on independent component analysis (ICA). *J. Neurosci. Methods* **307**, 125–137 (2018).
21. A. J. Watrous, J. Fell, A. D. Ekstrom, N. Axmacher, More than spikes: Common oscillatory mechanisms for content specific neural representations during perception and memory. *Curr. Opin. Neurobiol.* **31**, 33–39 (2015).
22. R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
23. J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, P. Pietrini, Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
24. N. Kriegeskorte, M. Mur, P. A. Bandettini, Representational similarity analysis—Connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
25. L. Deuker, J. L. Bellmund, T. Navarro Schröder, C. F. Doeller, An event map of memory space in the hippocampus. *eLife* **5**, e16534 (2016).
26. A. J. H. Chanales, A. Oza, S. E. Favila, B. A. Kuhl, Overlap among spatial memories triggers repulsion of hippocampal representations. *Curr. Biol.* **27**, 2307–2317.e5 (2017).
27. M. L. Schlichting, J. A. Mumford, A. R. Preston, Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nat. Commun.* **6**, 8151 (2015).
28. N. Axmacher, C. E. Elger, J. Fell, Memory formation by refinement of neural representations: The inhibition hypothesis. *Behav. Brain Res.* **189**, 1–8 (2008).
29. N. Saito, H. Mushiake, K. Sakamoto, Y. Itoyama, J. Tanji, Representation of immediate and final behavioral goals in the monkey prefrontal cortex during an instructed delay period. *Cereb. Cortex* **15**, 1535–1546 (2005).
30. A. J. Watrous, J. Miller, S. E. Qasim, I. Fried, J. Jacobs, Phase-tuned neuronal firing encodes human contextual representations for navigational goals. *eLife* **7**, e32554 (2018).
31. B. R. Moore, A modification of the rayleigh test for vector data. *Biometrika* **67**, 175–180 (1980).
32. M. W. Jones, M. A. Wilson, Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS Biol.* **3**, e402 (2005).
33. J.-P. Lachaux, E. Rodriguez, J. Martinerie, F. J. Varela, Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* **8**, 194–208 (1999).
34. P. Ravassard, A. Kees, B. Willers, D. Ho, D. A. Aharoni, J. Cushman, Z. M. Aghajan, M. R. Mehta, Multisensory control of hippocampal spatiotemporal selectivity. *Science* **340**, 1342–1346 (2013).

35. L. K. Vass, M. S. Copara, M. Seyal, K. Shahlaie, S. T. Farias, P. Y. Shen, A. D. Ekstrom, Oscillations go the distance: Low-frequency human hippocampal oscillations code spatial distance in the absence of sensory cues during teleportation. *Neuron* **89**, 1180–1186 (2016).
36. S. A. Lee, J. F. Miller, A. J. Watrous, M. R. Sperling, A. Sharan, G. A. Worrell, B. M. Berry, J. P. Aronson, K. A. Davis, R. E. Gross, B. Lega, S. Sheth, S. R. Das, J. M. Stein, R. Gorniak, D. S. Rizzuto, J. Jacobs, Electrophysiological signatures of spatial boundaries in the human subiculum. *J. Neurosci.* **38**, 3265–3272 (2018).
37. S. Maidenbaum, J. Miller, J. M. Stein, J. Jacobs, Grid-like hexadirectional modulation of human entorhinal theta oscillations. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 10798–10803 (2018).
38. D. Chen, L. Kunz, W. Wang, H. Zhang, W.-X. Wang, A. Schulze-Bonhage, P. C. Reinacher, W. Zhou, S. Liang, N. Axmacher, L. Wang, Hexadirectional modulation of theta power in human entorhinal cortex during spatial navigation. *Curr. Biol.* **28**, 3310–3315.e4 (2018).
39. G. Buzsáki, E. I. Moser, Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat. Neurosci.* **16**, 130–138 (2013).
40. B. C. Lega, J. Jacobs, M. Kahana, Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus* **22**, 748–761 (2012).
41. A. C. Heusser, D. Poeppel, Y. Ezzyat, L. Davachi, Episodic sequence memory is supported by a theta-gamma phase code. *Nat. Neurosci.* **19**, 1374–1380 (2016).
42. J. Vosskuhl, R. J. Huster, C. S. Herrmann, Increase in short-term memory capacity induced by down-regulating individual theta frequency via transcranial alternating current stimulation. *Front. Hum. Neurosci.* **9**, 257 (2015).
43. M. P. Noonan, N. Kolling, M. E. Walton, M. F. S. Rushworth, Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. *Eur. J. Neurosci.* **35**, 997–1010 (2012).
44. M. P. Noonan, M. E. Walton, T. E. J. Behrens, J. Sallet, M. J. Buckley, M. F. S. Rushworth, Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20547–20552 (2010).
45. M. P. Noonan, R. B. Mars, M. F. S. Rushworth, Distinct roles of three frontal cortical areas in reward-guided behavior. *J. Neurosci.* **31**, 14399–14412 (2011).
46. G. Buzsáki, Theta oscillations in the hippocampus. *Neuron* **33**, 325–340 (2002).
47. C. Qin, Z. Tan, Y. Pan, Y. Li, L. Wang, L. Ren, W. Zhou, L. Wang, Automatic and precise localization and cortical labeling of subdural and depth intracranial electrodes. *Front. Neuroinform.* **11**, 10 (2017).
48. R. M. Cichy, D. Pantazis, A. Oliva, Resolving human object recognition in space and time. *Nat. Neurosci.* **17**, 455–462 (2014).
49. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Proceedings of the 25th International Conference on Neural Information Processing Systems* (Neural Information Processing Systems Foundation, 2012), pp. 1097–1105.
50. M. X. Cohen, Fluctuations in oscillation frequency control spike timing and coordinate neural networks. *J. Neurosci.* **34**, 8988–8998 (2014).

Acknowledgments: We are grateful to all patients for participating in our study. We thank R. Oostenveld and A. J. Watrous for comments on the manuscript. We thank R. Heinen for estimating visual similarity of cue objects via a DNN. **Funding:** L.K., D.L.-P., and A.S.-B. were supported by the BMBF (01GQ1705A), NSF grant BCS-1724243, NIH grant 563386, and the BrainLinks-BrainTools Cluster of Excellence funded by the German Research Foundation (DFG; EXC 1086). L.W. was supported by the Strategic Priority Research Program of Chinese Academy of Science (XDB32010300), the Beijing Municipal Science and Technology Commission (Z171100000117014), CAS Interdisciplinary Innovation Team (JCTD-2018-07), and the Natural Science Foundation of China (81422024, 31771255). P.C.R. received support from the Else Kröner-Fresenius-Stiftung (Bad Homburg, Germany), the German Ministry for Economic Affairs and Energy (Berlin, Germany), and the Medical Faculty of the University of Freiburg (Freiburg, Germany). P.G. was supported by a research grant offered by the "Epilepsie-Akademie Berlin-Bethel," which is sponsored by the von Bodelschwingh Foundation Bethel, Bielefeld, Germany. N.A. received funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 316803389 – SFB 1280 as well as via Projektnummer 122679504 – SFB 874. **Author contributions:** The experiment was designed by L.K., L.W., A.Br., M.D., A.Bi., T.N.S., A.S.-B., and N.A. Epilepsy patients were implanted by P.C.R., V.A.C., W.Z., and S.L. Data were collected by L.K., L.W., D.L.-P., H.Z., A.Br., D.C., W.-X.W., P.G., and C.G.B. Analyses were performed by L.K. L.K. and N.A. wrote the manuscript. All authors provided essential feedback on the manuscript at different stages, contributed substantially to, and approved the final version of the manuscript. **Competing interests:** P.C.R. received travel, accommodations, meeting expenses, and/or speaking fees from Boston Scientific (Natick, MA, USA), Brainlab (Munich, Germany), and Inomed (Emmendingen, Germany). V.A.C. received grants for IITs from Boston Scientific (CA, USA) and Medtronic (USA) and travel grants and honoraria from Boston Scientific and Brainlab (Munich, Germany) and is a scientific advisor for CorTec (Freiburg, Germany). P.G. received honoraria for a speaking engagement from Eisai (Frankfurt, Germany). C.G.B. gave scientific advice to UCB (Monheim, Germany); obtained honoraria for speaking engagements from Eisai (Frankfurt, Germany), UCB (Monheim, Germany), Desitin (Hamburg, Germany), Biogen (Ismaning, Germany), and EUROIMMUN (Lübeck, Germany); received research support from Deutsche Forschungsgemeinschaft (Bonn, Germany), Gerd-Altenhof-Stiftung (Deutsches Stiftungs-Zentrum, Essen, Germany), diamed (Köln, Germany), and Fresenius Medical Care (Bad Homburg, Germany); and is a consultant to Laboratory Krone (Bad Salzuflen, Germany) regarding neural antibodies and therapeutic drug monitoring for antiepileptic drugs. The authors declare no other competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 23 October 2018
Accepted 24 May 2019
Published 3 July 2019
10.1126/sciadv.aav8192

Citation: L. Kunz, L. Wang, D. Lachner-Piza, H. Zhang, A. Brandt, M. Dümpelmann, P. C. Reinacher, V. A. Coenen, D. Chen, W.-X. Wang, W. Zhou, S. Liang, P. Grewe, C. G. Bien, A. Bierbrauer, T. Navarro Schröder, A. Schulze-Bonhage, N. Axmacher, Hippocampal theta phases organize the reactivation of large-scale electrophysiological representations during goal-directed navigation. *Sci. Adv.* **5**, eaav8192 (2019).

Hippocampal theta phases organize the reactivation of large-scale electrophysiological representations during goal-directed navigation

Lukas Kunz, Liang Wang, Daniel Lachner-Piza, Hui Zhang, Armin Brandt, Matthias Dümpelmann, Peter C. Reinacher, Volker A. Coenen, Dong Chen, Wen-Xu Wang, Wenjing Zhou, Shuli Liang, Philip Grewe, Christian G. Bien, Anne Bierbrauer, Tobias Navarro Schröder, Andreas Schulze-Bonhage and Nikolai Axmacher

Sci Adv **5** (7), eaav8192.
DOI: 10.1126/sciadv.aav8192