The Interplay of Attention and Conscious Perception: Evidence from Human Intracerebral Recordings and Computational Modeling

Jianghao Liu[1,2,*], Dimitri J. Bayle[3], Alfredo Spagna[1,4], Jacobo D. Sitt[1], Alexia Bourgeois[5], Katia Lehongre[6], Sara Fernandez-Vidal[6], Vincent Navarro[6,7,8], Claude Adam[6,7,8], Virginie Lambrecq[6,7,8], Tal Seidel Malkinson[1,†], Paolo Bartolomeo[1,†,*]

†These authors jointly supervised this work.
*Corresponding authors: jianghao.liu@icm-institute.org and paolo.bartolomeo@icm-institute.org

[1] Sorbonne Université, Inserm U 1127, CNRS UMR 7225, Paris Brain Institute, ICM, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France
[2] Dassault Systèmes, Vélizy-Villacoublay, France
[3] Licae Lab, Université Paris Nanterre, Nanterre, France
[4] Department of Psychology, Columbia University in the City of New York, NY, 10027, USA
[5] Laboratory of Cognitive Neurorehabilitation, Faculty of Medicine, University of Geneva; 1206 Geneva, Switzerland
[6] CENIR - Centre de Neuro-Imagerie de Recherche, Paris Brain Institute, ICM, Hôpital de la Pitié-Salpêtrière; 75013 Paris, France
[7] Service de neurologie 1, Hôpital de la Pitié-Salpêtrière; 75013 Paris, France
[8] Service de Neurophysiologie Clinique, Hôpital de la Pitié-Salpêtrière; F-75013, Paris, France

Abstract

Competing theories of consciousness disagree on the role of frontoparietal networks. Here we recorded neural activity from 727 intracerebral contacts in 13 epileptic patients, while they detected near-threshold targets preceded by supra-threshold exogenous cues. Trajectory *k*-means clustering revealed three patterns: (1) Validly cued seen targets elicit sustained activity in right-hemisphere fronto-temporal regions, connected by superior longitudinal fasciculus (SLF) II/III, and late accumulation activity in bilateral dorsolateral prefrontal cortex and in right-hemisphere orbitofrontal cortex (SLF I/III). (2) Invalidly cued seen targets elicit early, sustained activity in the right-hemisphere reorienting network (SLF III). (3) Seen targets elicit late sustained left dorsolateral prefrontal activity. Task modeling with recurrent neural networks supported the causal contribution of these networks to conscious perception, and elucidated their interplay through excitatory-inhibitory mechanisms. We provide new, compelling evidence on the role of hemisphere-asymmetric frontoparietal networks associated with attentional enhancement and reorienting in conscious perception.

Introduction

Rival theories of consciousness make different predictions on the role of fronto-parietal (FP) networks in conscious perception[1]. The recurrent processing theory[2] holds that conscious experience emerges from reverberating activity in sensory areas. The integrated information theory[3] postulates that conscious information is integrated in a temporo-parietal-occipital "hot zone", rather than in a FP network devoted to task monitoring and reporting[4]. According to both theories, the prefrontal cortex (PFC) is not essential for experience itself, but contributes to post-conscious cognitive processing, such as motor planning or verbal report[5]. Others claim that conscious processing is shaped by second-order metacognitive representations generated in the PFC[6]. The global neuronal workspace theory (GNWT) holds that a non-linear network ignition amplifies and sustains a global broadcasting of neural information, with an important role for dorsolateral PFC and inferior parietal cortex in this process[7]. According to GNWT, near-threshold stimuli must be attended to receive top-down amplification and attain conscious processing[8].

However, attention is a heterogeneous group of neurocognitive processes[9,10]. For example, endogenous and exogenous spatial attention refer to distinct behavioral and neural dynamics[11,12]. The available evidence suggests that endogenous, or top-down, spatial attention has little role in supporting conscious perception[13-15], while exogenous, or stimulus-driven, attention seems to be a necessary, although not sufficient, condition for conscious perception[16,17]. In neurotypical participants, exogenous cues increase the conscious detection rate of near-threshold targets, when they appear near the spatial location of the upcoming target (so-called valid cues)[18-20]. The dorsal FP attentional network[21] is more active for consciously perceived targets than for unseen targets at attended locations[22]. Clinical evidence from neurological

patients with signs of spatial neglect[23] strongly suggests a specific role for right hemisphere FP networks in conscious processing. These patients display a systematic pattern of association between right-biased exogenous attention and unawareness of left-sided events[24].

Despite this converging evidence, the fine-scale spatiotemporal dynamics of the modulation of exogenous spatial cues on conscious perception remains unknown. To fill this gap, here we used intracerebral EEG (iEEG), a technique with unmatched spatiotemporal resolution in humans. We also employed white-matter tractography to collect anatomical evidence on the network architecture, and computational modeling to make specific predictions on the causal relationships between attention and consciousness. We addressed the following questions: how does activity in ventral and dorsal FP networks[25,26] in either hemisphere[27] interact with conscious perception during exogenous orienting of attention? What happens in the brain when reorienting of attention to perceive unattended targets? Is it possible to adjudicate between potential roles of PFC in attentional orienting, working memory or decision making?

Results

Behavioral results: Cue validity modulates target detection

We recorded neural activity from 727 intracerebral contacts in 13 patients receiving presurgical evaluation of drug-resistant epilepsy (age 34.7 ± 8.7 years; 7 women). Patients performed a near-threshold targets detection task[28] (Fig 1a), in which they attempted to detect left- or right-tilted, near-threshold Gabor patches (the targets), presented either left or right of fixation. The target was preceded by supra-threshold peripheral non predictive visual cues, which appeared either on the same side as the subsequent target (Valid cues) or on the opposite side (Invalid cues). All conditions (target side, cue validity) were randomly interleaved, with 20% of cue-only, “catch” trials where no target was presented. Participants had to discriminate the direction of tilting, and

subsequently report the presence or absence of the Gabors. They were informed that cues did not predict the location of the upcoming targets.

To see whether exogenous attention modulates conscious perception of targets in the discrimination task, we performed a two-way ANOVA with the factors of cue validity and conscious report on the percentage of accurate responses. As expected, participants were more accurate in discriminating target direction when the targets were reported to be perceived than when they went unseen (main consciousness effect: $F_{1,24}$ = 24.74, p < 0.001, η^2 = 0.315). Although the main validity effect did not reach significance ($F_{1,24}$ = 1.04, p = 0.31, η^2 = 0.013), it interacted with conscious detection, because validly cued targets were more likely to be reported than invalidly cued ones ($F_{1,24}$ = 4.77, p < 0.05, η^2 = 0.061, see Fig 1b; post hoc *t*-test: valid vs. invalid for seen targets, p < 0.05). Signal detection theory (SDT) analysis[29] showed that valid cues induced a higher detection rate (pairwise *t*-test, p < 0.05, Cohen's d = 0.58), and a more liberal response criterion (pairwise *t*-test, p < 0.01, Cohen's d = 1.03) than invalid cues (Fig. S1a). However, no significant difference emerged in sensitivity (pairwise *t*-test, p = 0.52).

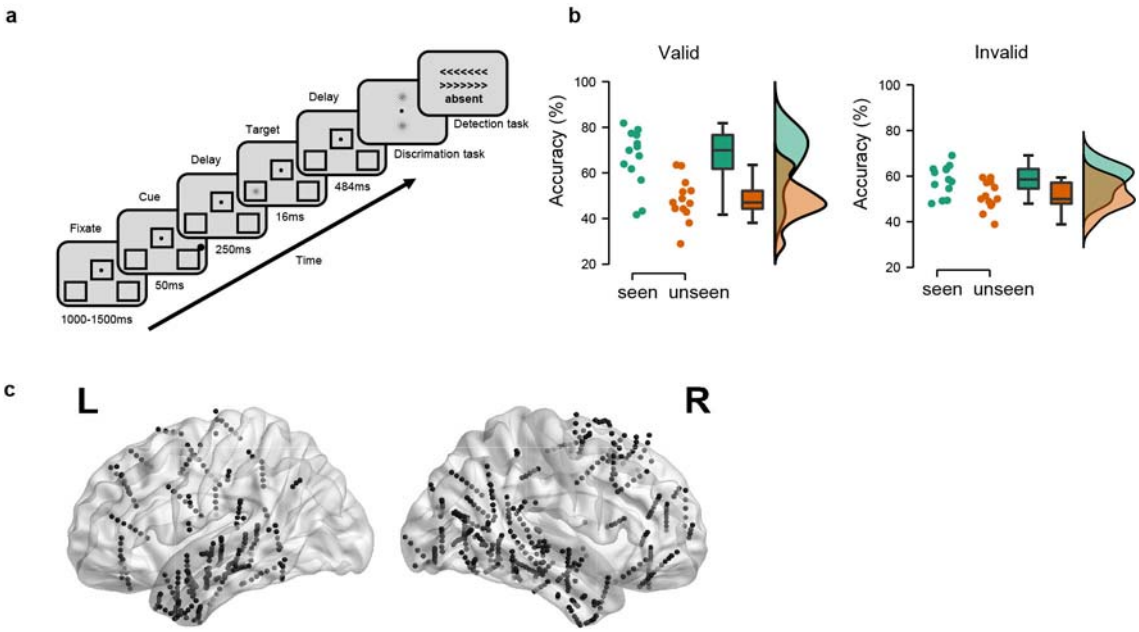

Fig 1.Near-threshold target detection task and human intracerebral recordings.
a. After 1000-1500ms of fixation, a peripheral non-informative dot cue occurred for 50ms, followed by a left-sided or right-sided near-threshold tilted Gabor target presented for 16ms. After a delay of 484ms, participants discriminated the direction of tilting and reported the presence or absence of the Gabor. In 20% of trials ("catch" trials), the Gabor target was absent.
b. Target discrimination accuracy. Dots represent individual performance. Chance level at 50%.
c. The 727 intracerebral contacts pooled from 13 epileptic patients in MNI space, see detailed localization label in Table S1.

Five neural clusters associate with consciously perceived targets

After discarding epileptic artifacts, there were 727 usable contacts with bipolar montage pooled across all participants (288 in the left hemisphere; 439 in the right hemisphere; Fig 1c, see Table S1 for details). For each contact, we extracted high-frequency broadband power (HFBB; 70-140 Hz), which is a proxy of neuronal population activity[30,31]. We then computed target-locked mean normalized HFBB across the eight experimental conditions (2x2x2 design: side of target [ipsilateral/contralateral] x cue condition [valid/invalid] x detection [seen/unseen]).

Next, we applied a trajectory *k*-means clustering method[32] to identify the main groups of contacts that carry cue validity and conscious report information (see Methods). This procedure

allowed us to group intracerebral contacts based on their temporal profile of neural activities across the experimental conditions. For each contact, we computed the temporal trajectory in the eight-dimensional condition space, i.e. the path of each contact's HFBB power over time across all conditions (target side, cue validity, detection). Using *k*-means clustering, each trajectory was then assigned to the cluster with the nearest trajectory-centroid, by iteratively minimizing within-cluster Manhattan distances. A ten clusters solution reflected the highest average silhouette score serving as the evaluation criterion of the clustering results (Fig S2a).

We next explored how our experimental manipulation of attention and consciousness influenced the clusters' activity. For each cluster, we performed time-resolved 3-way ANOVAs in both the cue-target period (from -300ms to target onset) and the post-target period (from target onset to 500ms post-target) with the factors of Target side, Validity, and Conscious report. Five of the ten clusters showed a main effect of conscious report, with higher levels of activity for seen than for unseen targets (all $ps < 0.018$). The number of contacts in these clusters was stable (Fig 2a) and their cluster-level temporal profiles were similar across *k*-means solutions with varying numbers of clusters. We thus focused on these clusters for further analyses (see Table S1 for details about anatomical localization in each cluster).

We will first describe the main effect of conscious report in the cue-target and post-target periods for each of the five clusters (Fig 2b). The first cluster (hereafter: the Visual cluster, 42 contacts) showed an early transient post-target effect, with stronger activation for seen targets compared to unseen ones (90-350ms and 380-430ms, all $ps < 0.003$). This cluster mainly consisted of contacts in the right posterior temporoparietal areas (there was no available electrode in the homolog areas of the left hemisphere). The second cluster (Sustained cluster, 148 contacts) showed an effect of consciousness both in early cue-elicited (-140 to -90ms before

target onset, all *ps* < 0.016) and in later, target-related sustained neural activity (160-200ms, 240-300ms, 340-430ms, and 450-500ms post target, all *ps* < 0.003). The Sustained cluster's contacts were mainly located in the bilateral temporal cortex, the right angular gyrus (AG), and the right PFC, around the central portion of the right superior frontal gyrus. The activity of the third cluster (Late accumulation cluster, 67 contacts) in the post-target period increased over time for seen targets (300-500ms, all *ps* < 0.005), but there was no significant conscious report effect in the cue-target period. Most of the contacts in the Late accumulation cluster were located in the bilateral PFC, around the left inferior frontal gyrus (IFG), the right orbitofrontal cortex (OFC) and the caudal portion of the right superior frontal gyrus. A fourth cluster (19 contacts) identified fast activity in the cue-target period (-250 to -220ms and -130 to -90ms before target onset, all *ps* < 0.015), as well as early in the post-target period (170-230ms, all *ps* < 0.001), that resulted in "seen" reports. This cluster was localized in the right hemisphere temporoparietal junction (TPJ) / IFG. A fifth cluster (38 contacts) showed a late sustained neural activity selective for seen targets (310-450ms, all *ps* < 0.018). This cluster contains contacts from the left posterior portion of dorsolateral PFC, around the left frontal eye field, and the bilateral posterior temporal area.

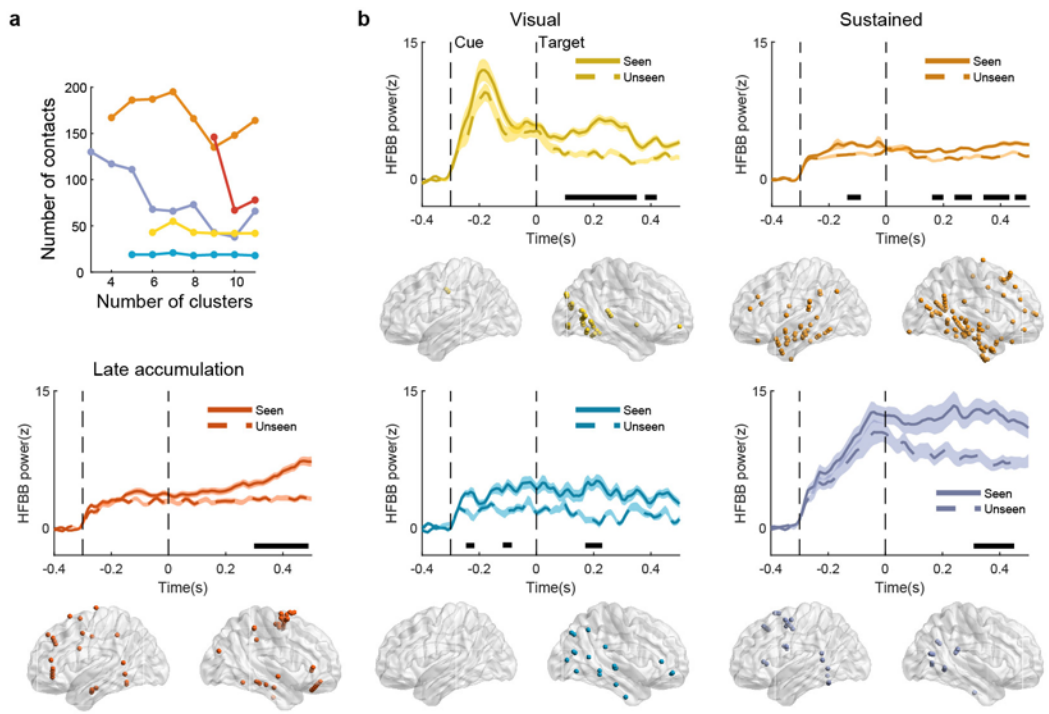

Fig 2. Trajectory k-means clustering revealed five clusters showing a conscious report effect.
a. Cluster stability. The change of the number of contacts across *k*-means solutions.
b. Comparison of target-locked neural activity for seen and unseen trials in each cluster. Black horizontal bars underneath the traces indicate $p < 0.05$, Holm-Bonferroni corrected. Shading is ± SEM across electrodes. Note that the Sustained cluster and the fourth (in blue) cluster also showed significant consciousness effect for the target during the cueing period. The localization of contacts in each cluster was visualized.

The Visual cluster also showed an interaction between target side and cue validity in the cue-target period (60-210ms after cue onset, all *ps* < 0.006, Fig S2d), with higher neural activity for contralateral than for ipsilateral cues. However, there was no significant Target side effect in the post-target period, perhaps because of the low, near-threshold intensity of the targets. No significant three-way interaction emerged (all Fs < 8.17, *ps* > 0.20).

The interaction between exogenous attention and conscious report reveals three neural patterns

We then explored how exogenous attention affects conscious reports of the targets in the five clusters, by examining the interaction between Cue validity and Conscious report (valid/invalid x

seen/unseen) in the above mentioned time-resolved ANOVA. This analysis revealed 3 distinct patterns of neural activity (Fig. 3a): 1) The first pattern encompassed three out of the five clusters and showed enhanced conscious report effect for the validly cued targets compared to invalidly cued targets (Visual cluster: 190-220ms, all *ps* < 0.03; Sustained cluster: 270-330ms, all *ps* < 0.002; Late accumulation cluster: 360-430ms, all *ps* < 0.006), mirroring the behavioral interaction. The amplitude of this attention-related enhancement did not differ across the three clusters (three-way ANOVA with the factors of cluster x validity x conscious report, all *ps* > 0.30), possibly due to faint target contrasts. A further explorative analysis comparing this enhancement, by Cohen's d values, around the time points where the interaction was significant, showed an increasing effect size gradient from the Visual, to Sustained, and to Late accumulation clusters (Fig S3d, one-way ANOVA: $F_{2,42}$ = 3.83, *ps* < 0.05, η^2 = 0.15, linear polynomial contrast, *p* < 0.05, t = 2.48). 2) The fourth cluster showed a reversed pattern, with higher activity for seen targets after invalid cues than for seen targets after valid cues, early after target onset (160-190ms, all *ps* < 0.002). This cluster (henceforth the Reorienting cluster) was likely to reflect reorienting of attention from the invalid cue to the opposite target. Activity in this cluster was sustained after target onset, then showed a transient peak around 180ms only for invalid seen targets; however, for seen targets after valid cues, activity decreased once targets appeared. 3) The fifth cluster showed a late sustained activity for seen targets, independent of cue validity (Consciousness cluster). The temporal components of the three neural patterns in the interaction between exogenous attention and consciousness can be visualized in a two-dimensional t-distributed stochastic neighbor embedding (t-SNE) decomposition (see Fig S3a).

We then sought to understand the functional roles of the clusters by relating the neural activities to the behavioral responses. In each cluster, we divided the trials (pooled across

conditions) into 20 quantiles according to their response time (RT) in the discrimination task (see Methods; note that in the discrimination task patients had to wait for the onset of a response screen in order to respond). We tested the relation of the neural activities in RT-bins using time-resolved one-way repeated measures ANOVA, in a time window from target onset to 1,000ms, to avoid the influence of neural activity in the subsequent trial. The Late accumulation cluster showed target-locked sustained neural activity in trials evoking slower RTs (340-1,000ms, all *ps*<0.003; see FigS3b), but not in trials with faster RTs . The other clusters had no or only transient (less than 300ms) RT effects for targets. We then compared the neural activity of the 10 fastest RT bins with those of the 10 slowest RT bins. The Late accumulation cluster showed greater accumulating and sustained activity for slower RTs responses than faster ones (Fig. S3c). Interestingly, the Visual cluster showed a stimulus-locked higher neural activity for faster RTs, perhaps because faster responses resulted from stronger evidence for target presence. No significant effect emerged in other clusters. Next, we visualized the neural activity across RT-bins over a longer time window, which contains the neural information until the button press (target onset to 5,000ms, see Fig. 3b). Here, activity might be affected by the subsequent trial; hence, we report only observational findings without statistical testing. We observed that the transient response elicited in the Visual cluster was locked to visual modulations (i.e., the appearance of cue, target, screen, and switching of the display after the discrimination task). The Late accumulation cluster showed accumulated and sustained activity until report. The Reorienting cluster shows a late transient response. The Consciousness cluster elicited sustained neural activity, which was locked to the visual percept and was not associated with the time of conscious report. To test whether these activities were related to motor preparation, we computed the beta band power (16 - 28 Hz), which typically decrease with motor planning[33]. There was no

sign of decreasing beta activity in the Late accumulation cluster and in the Consciousness cluster. This result suggests that motor preparation had little role in eliciting these clusters (Fig. S3e).

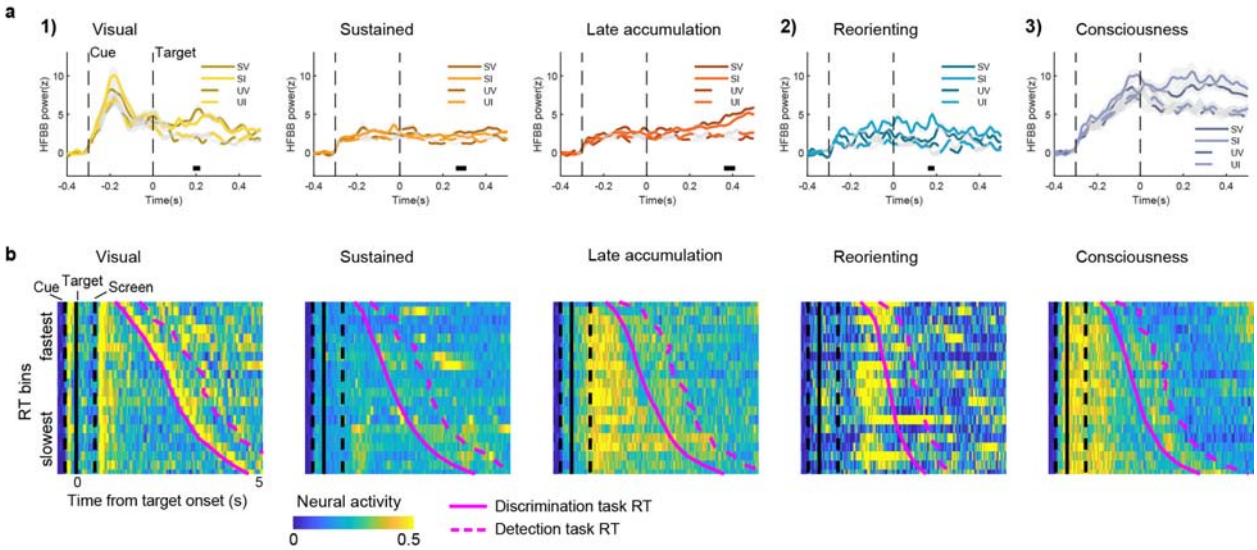

Fig 3. Cluster interaction of exogenous attention with conscious perception and functional modulation.
a. Three neural patterns of the interaction: 1) three clusters (Visual, Sustained, Late accumulation) showed enhanced conscious report effect for the validly cued targets. 2) a further, Reorienting cluster showed early sustained neural activity for invalidly cued targets. 3) a late Consciousness cluster differentiated seen from unseen targets independent of attention. SV: seen valid; SI: seen invalid; UV: unseen valid; UI: unseen invalid. Black horizontal bar for all *ps* < 0.05, Holm-Bonferroni corrected. Gray shading is ± SEM across electrodes.
b. Visual and RT modulation of target-locked neural activity. RT bins were sorted according to their mean RT from fastest to slowest, with neural activity pooled across contacts in each cluster. Magenta full curve shows mean discrimination task RT, followed by dashed magenta curve for mean detection task RT. Neural activity in the Visual cluster synchronized with visual stimuli (cue, response screen and screen switching after discrimination task). Late accumulation cluster showed sustained neural activity until the response. Consciousness cluster exhibited sustained neural activity locked to the visual percept, but not to the report.

FP white matter tracts connecting contacts in the right hemisphere

To specify the anatomical connections between contacts within each cluster, we performed white matter tractography analysis paired with probability maps in 176 healthy individuals from the Human Connectome database[34]. Fig. 4 displays the white matter tracts connecting frontal

contacts and parietal contacts within each cluster. This result suggests that our unsupervised cluster analysis mapped on existing anatomo-functional networks[27]. We examined the frontal association tracts described in Ref.[35], and found that our frontal and parietal contacts were connected by branches of the superior longitudinal fasciculus (threshold-free cluster enhancement-based non-parametric t-test, $p < 0.05$). Specifically, the Sustained cluster was mainly connected by the right superior longitudinal fasciculus (SLF) II (43.4%) and III (36.8%). The Late accumulation cluster was mainly linked by the right SLF I (39.4%) and III (25.1%). The Reorienting cluster was connected by right SLF III (84.2%). No statistically significant tracts emerged from the analysis of the Visual and the Consciousness cluster.

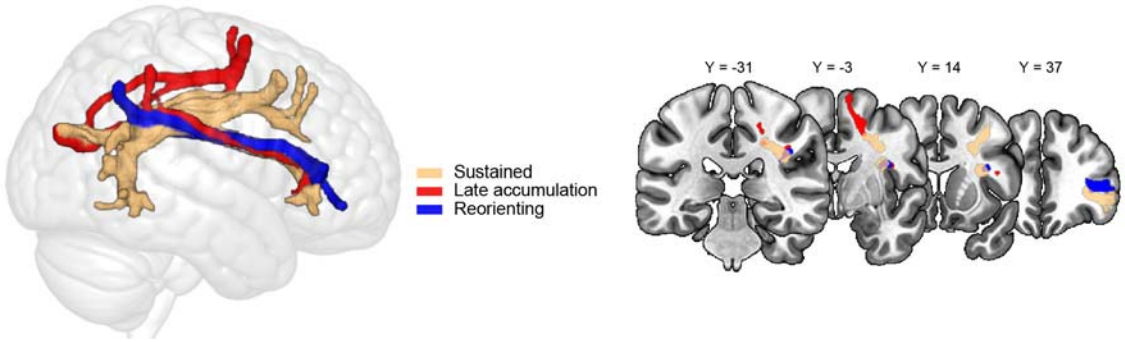

Fig 4. FP white matter tracts connecting contacts in the right hemisphere
Tractography t-maps, showing the significant white matter voxels (threshold-free cluster enhancement-based non-parametric *t*-test, $p < 0.05$), which connect frontal and parietal contacts within each cluster

Task modeling with recurrent neural network

To better understand the relation between activity dynamics and behavioral performance, we simulated the task with a recurrent neural network (RNN) model (Fig. 5a). We separately modeled left- and right-sided visual stimuli as two noisy signal inputs. The model had a single layer containing 50 units recurrently connected to one another, of which 80% were excitatory units and 20% were inhibitory units. The network was trained by back-propagation to produce two different outputs, one for each side, about present or absent targets. Similarly to the human

task, two RNN outputs were combined to measure a discrimination and a detection performance. The task conditions were also similar to the human task, and included valid trials, invalid trials, and 20% catch trials (see examples of stimuli inputs, hidden units dynamic and network outputs in Fig. 5b for valid trial and in Fig. S4a for invalid and catch trials).

The trained model displayed a detection psychometric curve, resembling typical human performance[36-38]. This curve differed from that characterizing an untrained model with random Gaussian connectivity weights (all *ps* < 0.001 for stimulus contrast above 0.03, see left panel of Fig.4c). Further, in the trained model, the improvement of target discrimination and detection with valid cues emerged only at sufficiently high stimulus contrast levels (Fig. 5c, middle and right panels). Consistent with this task performance pattern, a t-SNE visualization of RNN units' components showed a difference for valid *vs*. invalid trials only in intermediate or higher target contrasts, but not with lower contrast levels (Fig. S4b-c). Thus, all further analyses on RNN units were performed on the intermediate target contrast levels, corresponding to near-threshold targets in the human task.

To identify the temporal patterns of activity of the RNN hidden units, we used our trajectory *k*-means clustering method. Similar to the human results, the clustering analysis with silhouette evaluation resulted in five stable clusters (Fig. S4d) with different temporal trajectories. Four of these clusters showed stronger unit activity for seen targets than for unseen targets (Fig. 5d). Again similar to the human neural data, after an early, transient activity in a Visual cluster (8 units, 110-300ms, all *ps* < 0.017), there was sustained unit activity in a Sustained cluster (20 units, 140-750ms, all *ps* < 0.032), and late activity in a Late accumulation cluster (8 units, 300-730 ms, all *ps* < 0.038). Mirroring the model's task performance, the higher the target contrast, the greater the activity in these clusters. Importantly, these clusters showed

significantly enhanced activities for validly cued targets at sufficiently high stimulus contrast (Fig. 5e), akin to neural amplification. The fourth, Reorienting, cluster showed early transient activity related to "conscious" detection (11 units, 170-230ms, all *ps* < 0.019). A t-SNE visualization of unit activity of these clusters showed distinct unit component patterns for seen targets preceded by valid *vs*. invalid cues (Fig S5).

To examine potential similarities between RNN clusters and human neural clusters, we used a temporal trajectory correlation analysis. RNN unit activity in the Sustained, the Late accumulation, and the Reorienting clusters in the trained model were significantly more similar to the equivalent human neural clusters than to the clusters obtained from the untrained model (Fig. 5f, one-sided permutation test, all *ps* < 0.001). We examined how these model clusters are inter-connected, and the nature of their computation. In the trained model, we extracted input weights (sensory enhancement gain), connection weights between units, and output weights (report gain) (Fig. S4e). The model only constrained the total number of excitatory (E) and inhibitory (I) units. After RNN training, both excitatory and inhibitory units emerged in each cluster. After grouping separately excitatory and inhibitory units in each of the three clusters (a total of 6 groups), we computed a directed cluster connection graph by averaging unit-to-unit connection weights from one group ("pre-synaptic") to each of the other groups ("post-synaptic") and compared the resulting connections to random weights (Fig. 5g, one-sample *t*-test, all *ps* < 0.05, Holm-Bonferroni corrected). The stimuli input was mainly connected to the excitatory units in the Sustained cluster which was associated with the conscious processing-Sustained intracerebral cluster. The Late accumulation excitatory units were connected to the decision output, confirming its role in decision making. Notably, the Reorienting excitatory units also received a branch of stimuli input and showed strong excitatory connection to the inhibitory

units in the Sustained and in the Late accumulation clusters, reflecting the Reorienting cluster's role in early target monitoring and in executing inhibitory control over stimuli processing and decision making units.

Finally, we lesioned each of the RNN clusters to ascertain their causal contribution to task performance. For each cluster, we set all the unit weights to zero, and monitored the change in task performance (Fig. 5h). Lesion of either the RNN Sustained cluster or the Late accumulation cluster decreased the percentage of detected targets ("consciousness" units). Lesion of the Late accumulation cluster additionally impaired discrimination accuracy. Lesion of the Reorienting cluster led to a selective failure to detect and to discriminate invalidly cued targets ("reorienting" units). However, performance reverted to normal with very high target contrast levels, presumably because these contrast levels were sufficient to capture attention even without the contribution of the Reorienting cluster.

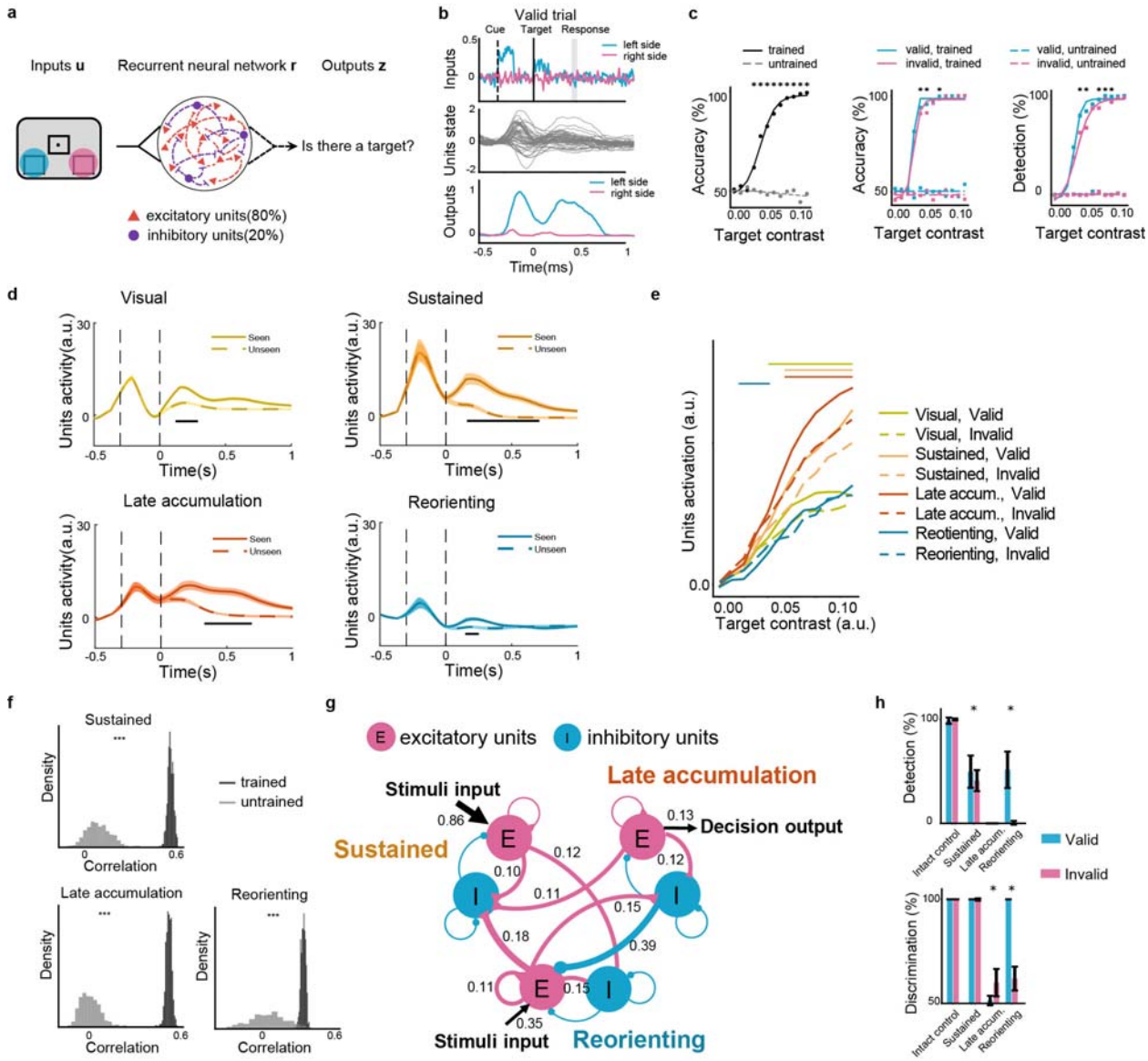

Fig 5. Task modeling with recurrent neural networks

a. Recurrent neural network model and task. Left- and right-sided visual stimuli were separately modeled as two noisy inputs. The model had a single layer containing 50 recurrent units. The network produced two outputs to decide whether a target was present on the left or on the right side.

b. Example of a trial showing the task structure and hidden unit dynamic as well as outputs from a trained model. After a fixation period, cue signal was presented on either the left or the right side, followed by a target signal presented on either side or absent. In a valid trial, cue signal and target signal were presented on the same side. Gray vertical line indicates the response window.

c. Task performance. Left panel: discrimination psychometric curve for all targets. Lines show the best-fitting logistic function to 12 target contrasts. Comparison of accuracy of the trained model to the untrained model by *t*-test, $*p < 0.001$, Bonferroni-corrected. Middle panel: discrimination psychometric curve. Comparison of discrimination accuracy for validly cued targets *versus* invalidly cued targets in trained model by *t*-test, $*p < 0.05$, Bonferroni-corrected. Right panel: detection psychometric curve for validly cued targets *versus* invalidly cued targets.

d. Trajectory *k*-means clustering of model unit activities. Four unit clusters showed distinct temporal trajectories in seen *versus* unseen trials for intermediate target contrast level. Black lines for all $*p < 0.05$, Holm-Bonferroni corrected. a.u.: arbitrary unit.
e. Comparison of unit activities for valid *versus* invalid trials in each cluster while increasing stimulus contrast. Late accum.: Late accumulation. Horizontal bar above the curves for all $*p < 0.05$, Holm-Bonferroni corrected. a.u.: arbitrary unit.
f. Histograms of the correlation coefficient between the trajectory of neural activity and unit activity. Permutation tests, $***p < 0.001$ for trained model versus untrained model.
g. Directed cluster connection graph for groups of excitatory and inhibitory units from three clusters. Magenta curves show excitatory connection and blue curves for inhibitory connection. Triangle or circle at the end of the connection line represents the destination. Numbers denote the weight of connection, with significance level controlled at $p < 0.05$ in one-sample *t*-test, Holm-Sidak corrected across groups of units.
h. Effect of cluster's lesions. Intact control: unlesioned model; Late accum.: Late accumulation; Reorienting. $*p < 0.05$ for the comparison of valid *versus* invalid trials by *t*-test.

Discussion

We combined human intracerebral EEG, white-matter tractography, and computational modeling to elucidate the fine-scale spatiotemporal dynamics of the modulation of exogenous spatial cues on conscious perception. Unsupervised temporal clustering revealed three patterns of neural activity in ventral and dorsal frontoparietal networks that supported the interaction between with exogenous orienting and conscious perception: (1) a Sustained cluster showing attention-enhanced sustained activity to validly cued targets; (2) a Late accumulation cluster with progressively increasing activity until report; (3) a Reorienting cluster showing an early, sustained response to invalidly cued targets. Altogether, our behavioral, neural, and modeling findings consistently demonstrate that exogenous attention, the process that makes external salient stimuli pop-out of a visual scene, modulates conscious perception. We also provide new, compelling evidence on the role of dorsolateral PFC, IFG and OFC in the human conscious processing of exogenously cued, near-threshold targets. These interactions between attention and conscious perception support and specify current models of consciousness, such as the gateway hypothesis[39], and the Global Neuronal Workspace hypothesis of consciousness. We

demonstrated distinct neural activity patterns that implemented attentional enhancement and reorienting in three FP networks in the right hemisphere, connected by branches of SLF.

At the behavioral level, valid cues increased both detection rate and discrimination accuracy of near-threshold peripheral target, a result that is in line with previous findings[17,19,20,40]. These behavioral effects were associated with neural dynamics in five clusters of iEEG contacts that fall into three distinct neural activity patterns. First, validly cued seen targets elicited stronger neural activity than invalidly cued ones. Additionally, after fast transient target-related visual activity, there was sustained activity in FP and temporal regions, and late accumulation activity in bilateral PFC. Second, invalidly cued seen targets elicited early, sustained activity in the right hemisphere TPJ and IFG (reorienting network)[25]. Third, independent of attention, seen, but not unseen, targets elicited late sustained activity in the left dorsolateral PFC. Furthermore, white matter tractography showed that the superior longitudinal fasciculi (SLF) II/III connected the FP areas shown to be implicated in sustained activity; SLF I/III connected the FP areas associated with late accumulation activity; SLF III connected the reorienting network, consistent with previous evidence[26]. Previous work on non-human primates showed that reported stimuli were associated with strong sustained activity in the frontal cortex[41]. However, this late PFC activity might reflect decision-making instead of conscious experience[5]. Our findings reconcile this debate by specifying the distinct involvement of three anatomically defined right-hemisphere FP networks in conscious perception.

The SLF I-III network showed a late accumulation and sustained activity until report, consistent with its known role in evidence integration and decision making[42]. However, activity around the mesio-frontal areas and supplementary motor areas might also reflect conscious

expectancy for target perception[43,44]. In line with early EEG studies[45], our results demonstrate that attention could enhance conscious expectancy through this FP network.

Cue-related activity in the SLF II network including the right superior frontal gyrus led to better conscious processing of subsequent targets. Attentional enhancement of conscious perception in the SLF II network provides a specific neural substrate for neural amplification suggested by the GNWT[8]. Importantly, transitory electrical inactivation of the SLF II in the human right hemisphere in a patient undergoing brain surgery provoked severe, if transient, inability to process left-sided stimuli, akin to signs of left spatial neglect[46]. Furthermore, damage to the right SLF II is the best anatomical predictor of the occurrence of neglect signs in stroke patients[23,47]. The present evidence specifies the temporal dynamics of the right hemisphere SLF II network, by demonstrating its role in attentional amplification of future targets.

Our results also underline the role of the right-hemisphere SLF III network in the conscious perception of targets preceded by an invalid cue at the “wrong” location. Previous neuroimaging evidence has shown the implication of this network in reorienting attention from the invalidly cued location to the targets occurring on the opposite side. Importantly, our time-resolved results specify that this activity occur earlier than previously thought, before target occurrence. This early activity must logically precede reorienting, and suggests an anticipatory ("lookout") activity for unexpected events such as invalidly cued targets. Consistent with this hypothesis, early SLF III network activity decreased when targets appeared in the validly cued location.

An additional cluster emerged in the opposite, left hemisphere, and was centered on the posterior portion of the left dorsolateral PFC. This cluster showed late sustained activity for

consciously perceived targets, independent of attention and motor response**.** In line with previous neuroimaging evidence[48], this neural activity might reflect the integration of sensory evidence and the formation of decision variables in post-orienting processes that are closely associated with conscious access mechanisms. Another possibility is that the left dorsolateral PFC biases perceptual decisions in conditions of uncertainty[49] such as near-threshold detection.

Our RNN model of the task displayed striking similarities with the human intracerebral data, and allowed us to make causal inferences on the functioning of the neural clusters we observed. Two distinct components of the attention/consciousness interactions, Attentional Enhancement and Reorienting, clearly emerged in the trained RNN. The Sustained network receives the majority of stimuli input and selectively enhances target-related information for conscious perception. The Reorienting network is engaged in the detection of unexpected events while executing inhibitory control, and, if necessary, stopping the ongoing stimulus processing to shift attention. The model implemented his inhibitory function by sending reorienting excitatory signaling to the local inhibitory neurons of other stimuli processing or decision making networks, a result that nicely dovetails the early post-target activation of the SLF III reorienting network in our neural data. The RNN lesion data support a causal contribution of the FP networks to conscious perception of near-threshold stimuli. Our model predictions are supported by neuroimaging evidence showing the involvement of right-hemisphere FP networks in attentional enhancement[50] and inhibitory control[27,51]. In line with the present results, in previous simulation work of physiological leftward bias (pseudoneglect) a similar excitatory influence of the right SLF III network on the dorsal attentional network also emerged[52]. Our results extended this prediction to an excitatory-inhibitory neuronal interaction for conscious perception. Strong inter-regional excitation balanced by local inhibition can enable reliable sensory signal propagation to

the PFC that leads to the "global ignition"[53] leading to conscious visual processing according to the GNWT[54]. More direct causal evidence of the interplay of these FP attentional networks for conscious perception comes from the finding that damage to the right dorsal PFC and decreased microstructural integrity of the SLF III impaired the conscious perception of near-threshold information[55].

Limitations of the present study include the limited cortical covering, which is determined solely by medical needs, and the pathological population included, although contacts with epileptic activity were excluded from analysis. Despite these considerations, the present evidence clarifies the role of distinct FP networks in the conscious perception of near-threshold stimuli, reconciling competing theories of consciousness concerning the role of PFC[1]. Our findings establish specific roles for the right-hemisphere SLF II network in the attentional enhancement of near-threshold targets, for the right-hemisphere SLF III network in perceiving previously unattended targets, and suggest a role of left-hemisphere dorsolateral PFC in perceptual decision. RNN modeling supported the causal contribution of these FP networks to conscious perception, and elucidated their interplay in shaping human conscious experience.

Supplementary Figures

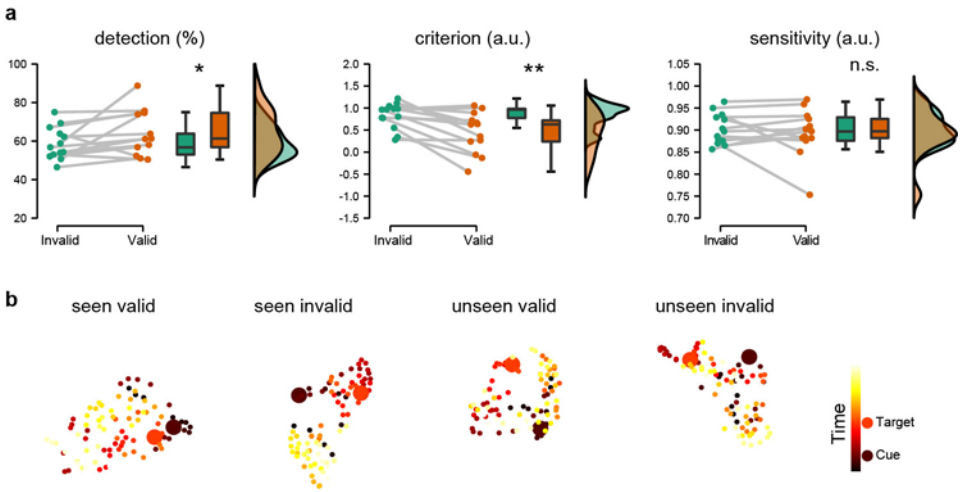

Fig S1. Signal detection theory analysis and visualization of neural activity components
a. comparison of detection rate, criterion and sensitivity in valid *versus* invalid trials. Dots represent individual performance. *$p < 0.05$; **$p < 0.01$; n.s.: not significant; a.u. arbitrary unit.
b. two-dimensional t-SNE visualization of neural activity components of all contacts. Color scales represent time points.

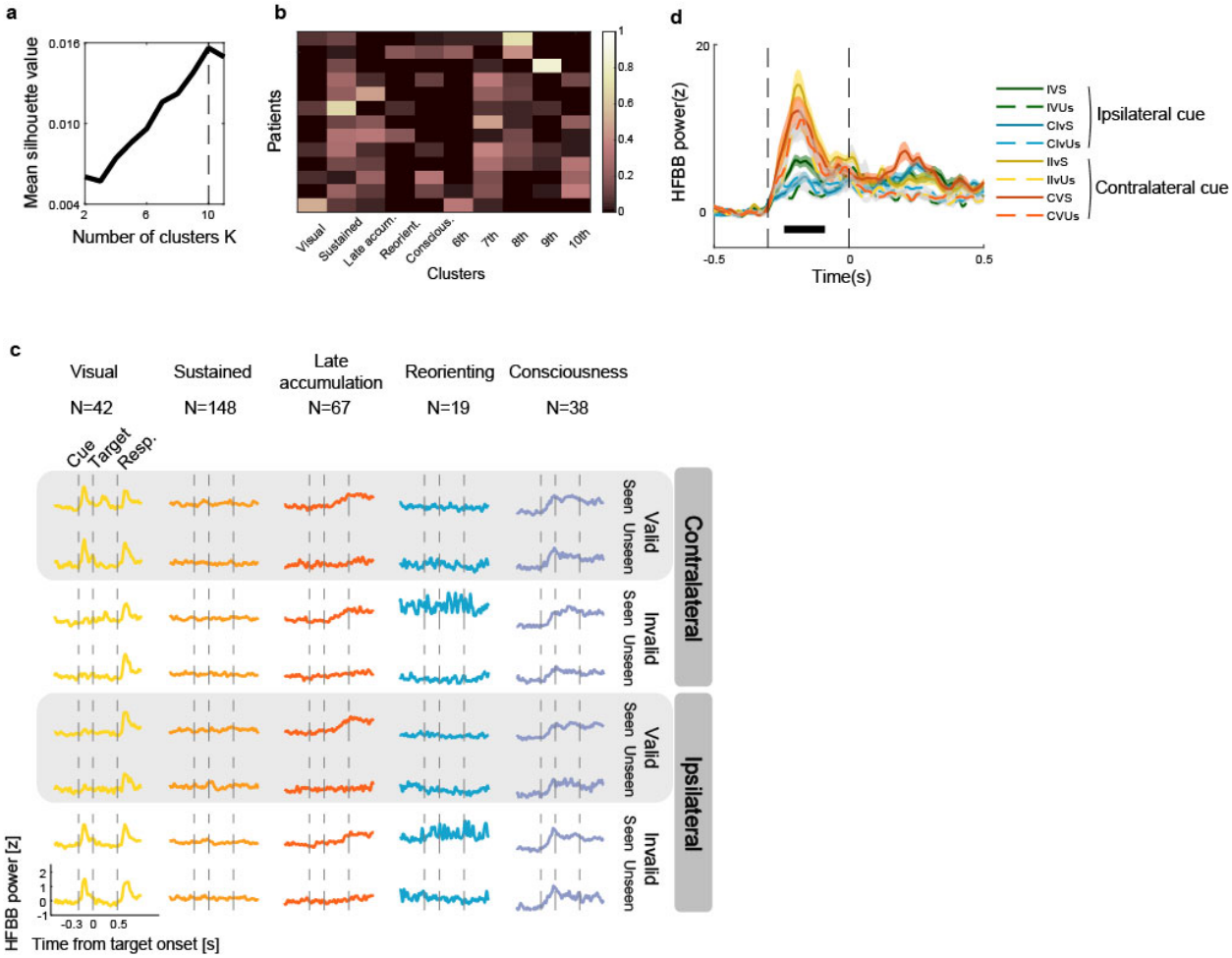

Fig S2. Clustering on the neural activities of intracerebral contacts

a. mean silhouette value of all contacts across k-means solutions.

b. partition of contacts by cluster for patients. Rows represent patients and color denotes the ratio of contacts in each cluster.

c. visualization of trajectory neural activities by conditions for k-means solution K = 10. Resp.:response screen. N: number of contacts in the cluster.

d. neural activity in the Visual cluster. During the cueing period (-300ms to 0), the Visual cluster showed a target side x validity interaction, with higher neural activity for contralateral cues than ipsilateral cues. Black horizontal bar for all $p<0.05$, Holm-Bonferroni corrected. I/C: Ipsilateral or Contralateral target; V/Iv: Valid or Invalid; S/Un: Seen or Unseen.

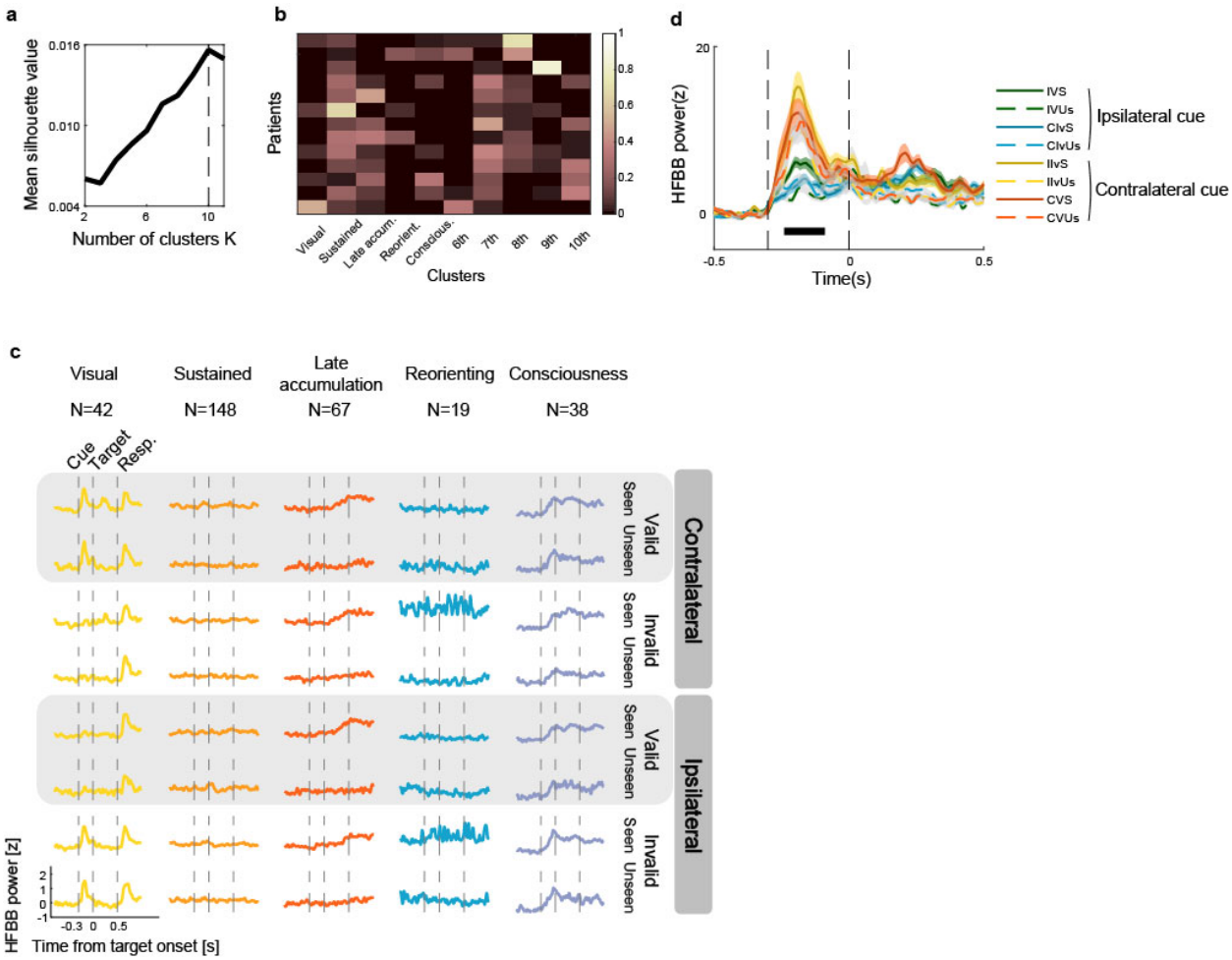

Fig S3. Intracerebral neural clusters neural activity

a. neural component visualization with t-SNE. Color represents time and dots represent neural activities by time points.

b. visual and RT modulation of target-locked neural activity in a short time window. In each cluster, the distribution of trial RT bins in all conditions was divided into 20 quantiles with equal probability. RT bins were sorted according to their mean discrimination task RT from fastest (in yellow) to slowest (in red), with neural activity pooled across contacts in each cluster. Time-resolved one-way ANOVA was performed on the neural activities of RT bins. Black horizontal bars indicate $p < 0.05$, Holm-Bonferroni corrected.

c. comparison of neural activities of the ten fastest and the ten slowest RT-bins by time-resolved *t*-test. Black horizontal bar for all $p < 0.05$, corrected by Holm-Bonferroni method.

d. linear contrast comparison of attentional enhancement effect across the Visual, Sustained and Late accumulation clusters. Cohen's d was derived from time-resolved *t*-test for the contrast (seen valid - unseen valid) - (seen invalid - unseen invalid) around the significant interaction time point in each cluster. Late accum.: Late accumulation. Dots represent Cohen's d values at time points.

e. beta band power. No decreased beta activity was observed.

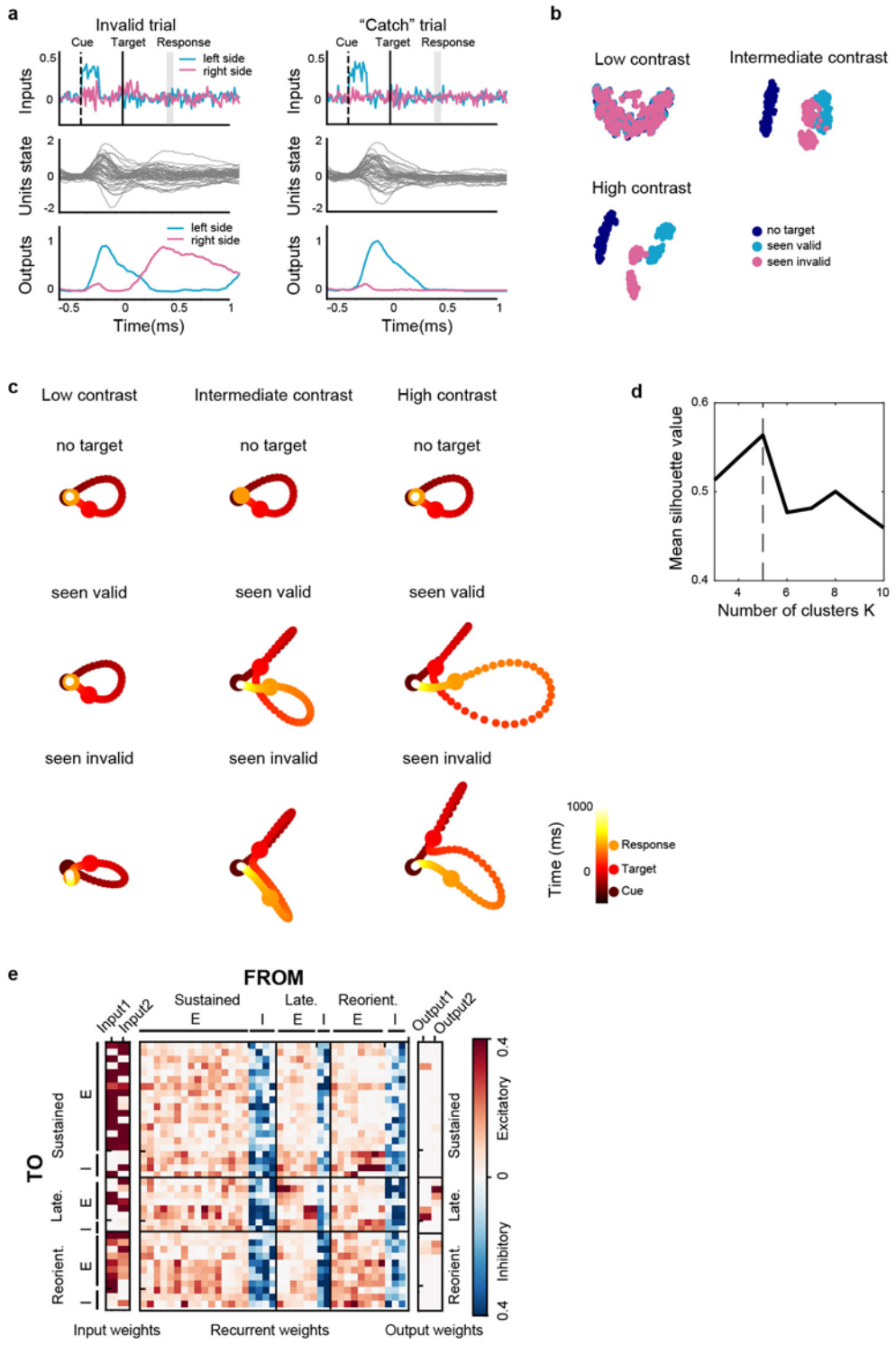

Fig S4. Task representation and unit activity in recurrent neural network model
a. example of task structure in invalid (left panel) and in target-absent "catch" trials (right panel).
b. two-dimensional t-SNE visualization of unit activities with all units, for low, intermediate and high target contrast levels, respectively. Dots represent trials.

c. two-dimensional t-SNE visualization of unit components with all units. Color scale represents time points.
d. mean silhouette values of all units across k-means solutions in clustering of RNN units.
e. directed connection weights of the task-optimized trained model. Left columns: input weights devoting enhancement gain; Middle columns: directed unit-to-unit connection weights; Right columns: output weights devoting report gain). Connections go from unit columns ("pre-synaptic") to unit rows ("post-synaptic"). Red represents excitatory weights and blue represents inhibitory weights. Late.: Late accumulation; Reorient.: Reorienting.Each cluster contains both excitatory (E) and inhibitory (I) units.

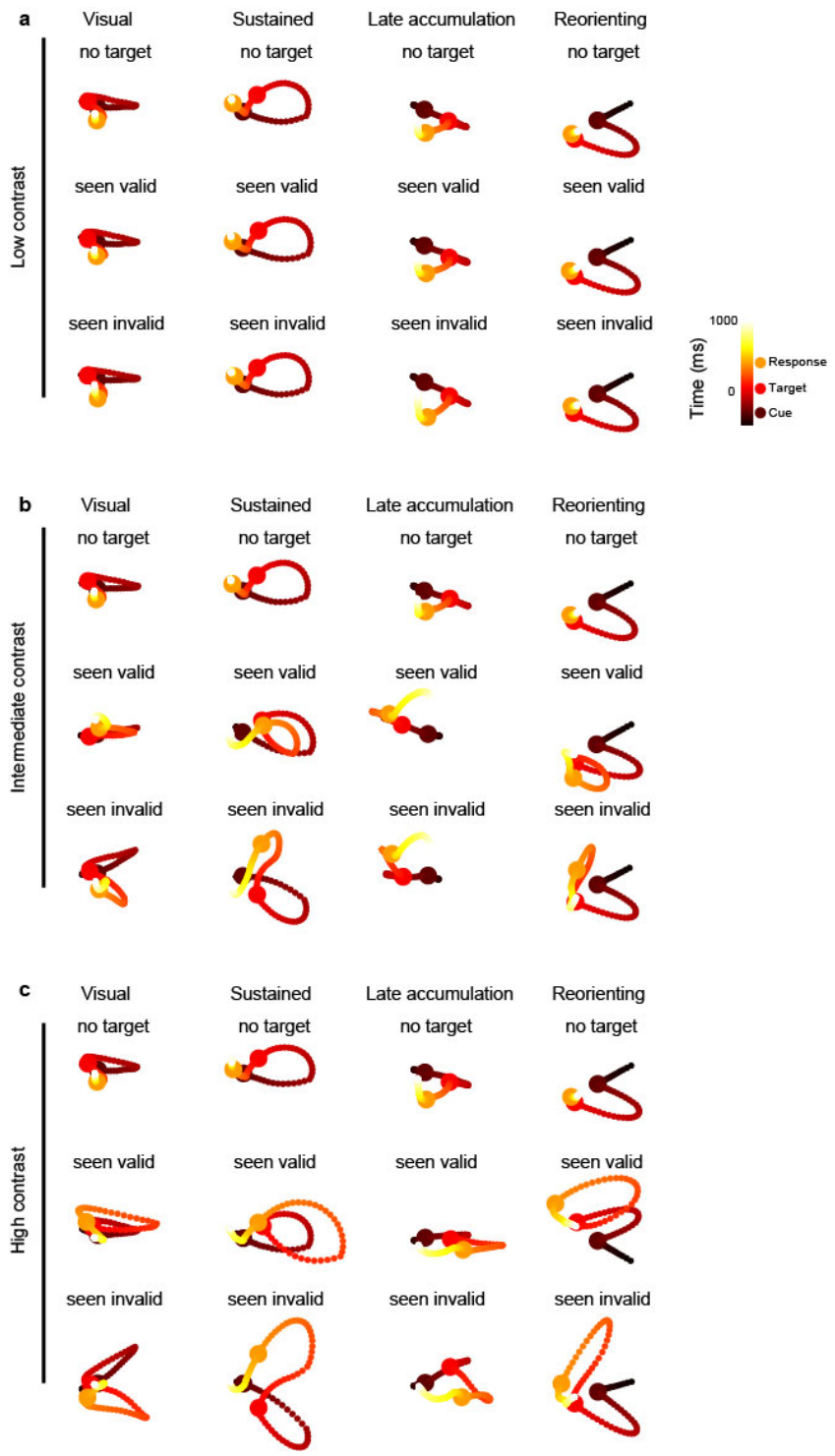

Fig S5. Unit components of unit clusters

Two-dimensional t-SNE visualization of unit components for low (a), intermediate (b) and high (c) target contrast levels by clusters. Color scales represent time points. Difference of unit component for validly and invalidly cued seen targets emerged when target contrast attained intermediate level, and increased for higher target contrasts.

Table S1. Electrodes localization according to the Desikan-Killiany-Tourville atlas.

Region name	Implanted Electrode N	Visual N	Sustained N	Late accumulation N	Reorienting N	Consciousness N
Banks superior temporal sulcus	4	0	1	0	0	0
Caudal anterior-cingulate cortex	11	0	2	0	0	0
Caudal middle frontal gyrus	20	0	2	3	0	2
Cuneus cortex	1	0	0	0	0	0
Entorhinal cortex	5	0	1	0	0	0
Fusiform gyrus Posterior	21	4	2	0	0	6
Fusiform gyrus Middle	16	2	3	1	0	2
Fusiform gyrus Anterior	4	0	3	0	0	0
Inferior parietal cortex	39	9	8	1	6	4
Inferior temporal gyrus Posterior	16	2	4	2	0	0
Inferior temporal gyrus Middle	17	0	5	1	1	0
Inferior temporal gyrus Anterior	22	0	10	0	0	0
Lateral occipital cortex	12	0	1	0	1	0
Lingual gyrus	15	0	2	0	0	0
Medial orbital frontal cortex	12	0	2	4	0	0
Middle temporal gyrus Posterior	61	13	15	4	3	2
Middle temporal gyrus Middle	30	0	13	1	1	0
Middle temporal gyrus Anterior	30	0	2	0	0	0
Parahippocampal gyrus	7	0	3	3	0	0
Paracentral lobule	5	0	1	0	0	0
Pars opercularis	10	0	0	1	0	2
Pars orbitalis	31	1	3	6	2	2
Pars triangularis	8	0	1	3	1	0
Pericalcarine cortex	1	0	0	0	0	0
Postcentral gyrus dorsal	5	0	0	0	0	0
Posterior-cingulate cortex	6	2	0	1	0	0
Precentral gyrus dorsal	15	0	0	3	0	6
Rostral anterior cingulate cortex	4	0	1	0	0	0
Rostral middle frontal gyrus	13	0	1	3	0	0
Superior frontal gyrus	73	0	11	18	0	5
Superior temporal gyrus Posterior	33	1	7	0	3	2
Superior temporal gyrus Middle	22	0	4	0	0	0
Superior temporal gyrus Anterior	25	0	6	0	0	0
Supramarginal gyrus	22	0	1	4	1	3
Temporal pole	19	0	10	0	0	0
Postcentral gyrus ventral	6	0	1	0	0	0
White matter	62	8	14	5	0	2
Hippocampus	24	0	8	3	0	0
Total	**727**	**42**	**148**	**67**	**19**	**38**
Frontal	88	1	24	42	3	18
Parietal	39	11	10	5	7	6

Methods

Participants and intracerebral recordings

We recruited sixteen patients who underwent presurgical evaluation of pharmaco-resistant focal epilepsy with intracerebral encephalography (iEEG) implantation, from the Department of Neurosurgery of the Hôpital Pitié-Salpêtrière, Paris, France. All participants had normal or corrected-to-normal vision and provided their written informed consent (CPP Paris VI, INSERM C11-16 (2012-2020); CPP INSERM C19-55). Three patients were excluded from the data analysis due to bad data quality (two patients had unreadable data and one patient had many response times inferior to 150ms), leaving totally thirteen included patients (age mean ± SD: 34.6 ± 8.7; 6M). Patients were implanted with 4-13 multilead stereotactic depth electrodes (AdTech®, Wisconsin) endowing 4-12 platinum contacts with a diameter of 1.12 mm and length of 2.41 mm, with nickel-chromium wiring. The distance between the centers of two contacts was 5 mm. Patients' electrode implantation was based exclusively on clinical criteria. In five patients the neural activity was recorded with a 128 channels clinical video-EEG recording system (SD LTM 64 BS, Micromed® S.p.A., Italy), sampling at 1024 Hz with a band-pass filter of 0.15 Hz to 250 Hz. In the other eight patients, the recording was implemented with a Neuralynx system (ATLAS, Neuralynx®, Inc., Bozeman, MO) which allowed recording up to 160 depth-EEG channels sampled at 4 kHz with 0.1Hz to 1000 Hz band-pass. The patient-dependent least active contact, preferably in the white matter, was selected as the reference electrode.

Experimental task

The stimuli presentation was controlled by E-Prime 2.0 software (SCR_009567) running on a laptop refreshing at 60 Hz. Three black boxes (4.9° long and 3.6° large) were arranged around a central fixation point, persisted for the whole duration of the trial, with 6° horizontally separating

the central box center from the peripheral boxes center. To maximize signal responses from early visual areas the two peripheral boxes were located in the lower visual field, 4° of visual angle under the central box. Fig. 1a illustrates the experimental procedure. Participants were instructed to fix their gaze at the central fixation throughout the test and respond as fast and accurately as possible. Following the appearance of the fixation and the three placeholder boxes for 1,000-1,500ms, peripheral cues consisting of a black dot (1° diameter) were presented for 50ms at the upper external corner of one of the two peripheral boxes. 300ms after the visual cue onset, a target stimulus presented for 16ms in one of the two peripheral boxes. The target stimuli were tilting Gabor patches with the spatial frequency of 5 cycles and the diameter of 2.5° visual angle, chosen among 12 equally spaced between 0 and 180°, excluding vertical and horizontal orientations. After a 484ms delay from the target offset, participants had to discriminate the direction of tilting between two possibilities distant by 30° (discrimination task), with forced-choice in the absence of targets. Participants subsequently reported the presence or absence of the Gabors (detection task), by selecting one of two opposing arrows, or the word 'absent' under the arrows. Both tasks lasted until participant's response or for a maximum of three seconds. Participants performed eight recording blocks, each consisted of 110 randomized trials including 88 target-present trials and 22 target-absent "catch" trials. Participants were informed that cues were non-informative, i.e. in target-present trials cues indicated the target location in 50% of trials (validly cued) and the opposite location on the remaining 50% of the trials (invalidly cued). Before the recording blocks, participants performed a target contrast calibration block in order to estimate the individual perceptual threshold contrast for 50% seen targets. The calibration consisted of two randomly interleaved, one-up one-down staircases, converging toward a detection rate of 50%.

Behavioral analysis

For each participant, we first excluded trials with response time (RT) faster than 150ms or exceeding three standard deviations. Statistical tests in behavioral analysis were performed using JASP software (version 0.16.0.0, https://jasp-stats.org/). We performed a two-way ANOVA with the factors of cue validity and conscious report on the discrimination accuracy. RTs analysis were not reported since participants had to wait for the response screen to give their responses.

Using a nonparametric measure[56], we conducted Signal Detection Theory (SDT) analysis[29] to evaluate the bias by the cue validity to participants' perceptual sensitivity. We computed the mean percentage of seen targets when the Gabor was presented (Hits) and when the Gabor was absent (false alarms; FA). The following formulas were used to calculate a': a' = 0.5 + (Hits - FAs) * (1+Hits-FAs) / [4 * Hits * (1-FAs)]; The *criterion (C)* summarizes the distance of the threshold relative to the noise distribution from the threshold of an ideal observer. A smaller value of *C* represents a more liberal threshold in target detection. We compared detection rate, sensitivity, and criterion between valid and invalid trials with paired sample *t*-test.

iEEG preprocessing

Spatial localization of the electrode has been recovered automatically using the Epiloc toolbox[57] developed by the STIM engineering platform in the Paris Brain Institute (https://icm-institute.org/en/cenir-stim) with co-registered pre-implantation 1.5T or 3T MRI scans and post-implantation CT and MRI scans. The normalization of MRI-pre, MRI-post and CT-post into the MNI space were applied using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/). Contact localization was automatically labeled referring to Desikan-Killiany-Tourville atlas parcellation[58] in the patient's native space, using Freesurfer image analysis suite (https://surfer.nmr.mgh.harvard.edu/) embedded in the Epiloc toolbox, followed by a manual verification and correction, if necessary.

We conducted the signal preprocessing using Matlab (R2018b, The MathWorks, Inc.) and FieldTrip toolbox[59]. Before analysis, all signals were down sampled to 512 Hz and all contacts were re-referenced to their adjacent neighbor contact on the same electrode, yielding a bipolar montage, in order to suppress non-local assemblies and ensure that iEEG signals could be considered as originating from a cortical volume centered within the two contacts. Coordinates of bipolar contacts were computed as the mean of the MNI coordinates of the two contacts composing the bipole. An initial visual inspection of continuous signals was performed to remove time segments showing transient epileptic or interictal activity. Contacts with excessive epileptic spikes or near suspected epileptic focus were also rejected. We extracted time courses from -1,300 to 1,200ms around target onset for trial epochs. A second artifact inspection was performed on the epoched data, where trials and contacts with excessive maximal signal, z-value, variance or kurtosis of the signal distribution were rejected. The data was also notch filtered at 50Hz as well as all harmonics corresponding to the frequency of alternating current. After signal preprocessing, 727 of the bipolar contacts were retained for further analysis. 288 contacts were in the left hemisphere and 439 contacts were in the right hemisphere. Fig. 1c and Table 1 show the localization of usable contacts referring to Desikan-Killiany-Tourville atlas parcellation[58]: 29 contacts (4%) from occipital, 332 (46%) from temporal, 78 (11%) from parietal, 202 (28%) from frontal, 24 (3%) subcortical and 62 (9%) in white matter.

We then adopted a pseudo-whole-brain analysis approach, by pooling contacts across all thirteen patients on a standardized brain in MNI space. High-frequency broadband (HFBB) power (70-140Hz), a proxy of spiking activity of the local neuronal ensemble[30,31], was extracted from each bipolar contact by wavelet time frequency transformation using the Morlet wavelets implemented in Fieldtrip (ft_freqanalysis), in fourteen equally spaced center frequency bands.

We kept high-frequency band power time courses from -800 to 1,000ms to target onset to discard the 1/f signal drop off at the edges. Baseline correction was applied on each trial by means of a z-score relative to the period in the 200ms prior to cue onset. Finally, HFBB powers were down-sampled to 100Hz for further analysis.

Temporal embedding visualization with t-SNE

We visualized neural activities and unit activities in a two-component space by a machine learning visualization approach, t-distributed stochastic neighbor embedding (t-SNE)[60]. In neural data, we averaged neural activities across contacts by conditions and time points, of which the resulting matrix served as input to compute temporal embedding. t-SNE was computed by an implementation in scikit learn 1.0.2 in Python 3. We adopted a perplexity of 30 and a learning rate of 100. The embedding was initiated with the PCA option and optimized upon 1000 iterations. The neural data in each condition was projected to a two-dimensional t-SNE embedding. Thus, the ensemble of time points formed a temporal trajectory of neural activities in a reduced manifold. A larger difference between two trajectory components represented a more distinct neural activity pattern between two conditions. In RNN simulation, we averaged unit activity across units by conditions and time points, and kept the above parameters in computing its t-SNE temporal embedding.

Trajectory k-means clustering

We applied a novel clustering approach based on *k*-means clustering to classify contacts by their temporal profiles[32], implemented through Matlab (R2018b, The MathWorks, Inc.). This data-driven approach was able to capture the prototypical patterns of neural dynamics that might be sensitive to cue validity and seen/unseen reports. We conducted clustering on all bipolar contacts. In each contact, we took the trajectories of the mean target-locked activity across an 8-

dimensional condition space (target side, cue validity and seen/unseen report). Activity across conditions was z-scored relative to the distribution of the trials' entire duration. Contacts were then iteratively partitioned (10000 iterations) into 2-12 clusters, in which each contact was assigned to the cluster with the nearest centroid trajectory. This was achieved by minimizing the sum of time-point-by-time-point Manhattan distances across conditions, to quantify trajectories similarity while preserving temporal order. Based on the silhouette evaluation (*silhouette* in Matlab), we adopted a ten clusters solution, which reflected the highest average silhouette score (Fig S2a).The partition of the 13 patients' contacts to clusters is shown in Fig S2b, demonstrating that clusters did not result from any single participant's trajectory activity, but rather reflected temporal patterns across many participants. In order to identify the consistency of clusters across different numbers of clusters K, we inspected clusters' trajectory profiles in each number of clusters. We plot the trajectories of five clusters that showed significant consciousness effects in K =10 cluster solutions (Fig S2c). The variation of the number of contacts in each cluster demonstrated the stability of the contacts of five selected clusters across k-means solutions (Fig 2a).

Consciousness and interaction-related neural activity

To explore how our experimental manipulation of attention and consciousness influenced the clusters' neural activity, we performed time-resolved three-way ANOVA tests with the factors of target side, cue validity and seen/unseen report. We tested on HFBB power in both cue-target period (from -300ms to target) and post-target period (from target onset to 500ms), across contacts on each cluster. For each contact, we averaged HFBB power across trials by conditions. Holm-Bonferroni correction was applied over all the time points for multiple comparisons. For clusters showing significant consciousness effect, post-hoc comparisons were performed on time

points where the Validity x Consciousness interaction was significant, with Holm-Bonferroni correction. Further, in order to compare the degree of attentional enhancement across clusters that showed higher neural activity for validly cued seen targets, we performed a linear contrast testing around time points where the clusters showed significant effects.

Response time (RT) modulation of neural activity

To understand the functional roles of neural clusters, we related the neural activity to the response time. In each cluster, we pooled trials from all conditions and contacts. We then sorted the trials and binned them into 20 quantiles according to their response time in the discrimination task. We compared RT-bins neural activity using time-resolved one-way ANOVA in a short time window (target onset to 1,000ms, to avoid the influence by the neural activity of the subsequent trial). To specifically test whether higher sustained neural activity was related to the slower response time, we then compared the neural activity of the 10 RT bins with slowest response to the neural activity of the 10 RT bins with fastest response. Holm-Bonferroni correction was applied over all the time points. We also visually compared the RT bins sorted neural activity in a longer time window (target onset to 5,000ms) in order to identify the neural activity patterns associated with visual modulation and task report in each cluster.

White matter tracts computing

We tracked the white matter tracts connecting parietal contacts with frontal contacts. We modeled each intracerebral contact by a 3 mm-sphere. In each cluster, we created two region-of-interests (ROIs), respectively consisting of frontal contacts and parietal contacts. The parcellation of contacts was referred to Desikan-Killiany-Tourville atlas, see details in iEEG preprocessing and Table S1 for contact numbers. We used a dataset that includes 176 preprocessed healthy individuals tractography acquired at 7 tesla by the Human Connectome

Project team[34]. We performed tracts filtering in TrackVis toolbox[61] to obtain tracts connecting frontal contacts and parietal contacts in frontal association tracts[35] (three branches of the superior longitudinal fasciculus; uncinate; long and anterior segments of the arcuate fasciculus; inferior fronto-occipital fasciculus). For each cluster, the tractography maps of the 176 healthy individuals were subsequently binarised and then smoothed with a three-dimensional Gaussian filter (full width at half-maximum was 5 mm, equivalent to a sigma of 2.123). To test the presence of tracts across individuals, we used a threshold-free cluster enhancement (TFCE)-based non-parametric test, with 1000 permutations (*"randomize"* function in FSL) and a height threshold of 0.95 to control significance level at $p<0.05$. We computed the volumetric ratio of the labeled tracts in each cluster with those of the standard atlas in BCBtoolkit[62] (http://toolkit.bcblab.com/), where we filtered the atlas probabilistic maps with a threshold of 80% to reduce their overlapping.

Task-optimized recurrent neural network model

Recurrent neural networks (RNNs) are networks in which neurons can send and receive feedback to and from each other. Therefore, the activity of neurons is affected not only by the current external stimulus, but also by the current state of the network[63], which makes RNNs ideally suited for computations that unfold over time such as holding the information of cue position or accumulating target-related evidence for decision making. If trained RNN accomplishes the behavioral task with a performance comparable to the human ones, the RNN hidden unit activities could provide unique insight about its computations in task representation. Moreover, RNN unit activities might also appropriately predict neural processing[64,65].

Our RNN model contained a single layer trained with mini-batch gradient descent learned by backpropagation. Before time discretization, the network activity r follows a continuous dynamical equation:

$$\tau \frac{dr}{dt} = -r + f(W_{rec}r + b_{rec} + W_{in}u + \sqrt{2\tau\sigma_{rec}^2}\xi)$$

$$r = f(x)$$

$$z = W_{out}r + b_{out}$$

where u , x and z denote the input, recurrent state, and output vectors, respectively. W_{in},W_{rec},W_{out} are the connection weight matrices of the input layer (a *2* x *N* matrix), the recurrent layer (*N* x *N*) and the output layer (*N* x *2*). b_{rec} and b_{out} are constant biases into the recurrent and output units. The network is time-discretized with positive activity. dt is the simulation time-step and τ is an intrinsic timescale of recurrent units which was set to 100ms. ξ denotes the independent Gaussian white noise processes with zero mean and unit variance, and σ_{rec} is the strength of the noise. f is a nonlinear transfer function, which was set as a rectified linear activation function (ReLU).

Similar to previous studies training RNNs to perform cognitive tasks[66-68], we abstracted the relevant visual stimuli properties from the patient task (Fig. 1a), rather than feeding the exact same visual inputs to the RNNs. Specifically, visual stimuli from the left and the right field of view were modeled as two separate noisy inputs (Fig. 5b). The magnitude of the background visual noise along the trial was set as Gaussian noise of the mean of zero and standard deviation (SD) of 0.05. The task began with a fixation period of 200ms followed by a visual cue randomly presented at either the left or the right side. A target, separated from the cue by a fixed delay

period of 300ms, was then appeared on the cued position (validly cued) or on the opposite side (invalidly cued) with equal probability, or absent in 20% of "catch" trials. After a second fixed delay, the network produced two outputs, each ranging from zero to one, to respond whether there was a target detected. Similar to the human task, RNN outputs were combined to calculate a discrimination and a detection performance. The discrimination accuracy is the ratio of correct response with force-choice in distinguishing target side when the target were presented. The detection rate equals the ratio of trials in which the network made the correct response about the target presence or absence.

We implemented the model training with PsychRNN[69], a toolbox backended by TensorFlow. We adopted the default setting of the package regarding the regularizers (i.e. penalties added to prevent over-fitting to the training data), weight initializer (i.e. by a glorot normal distribution) and the loss function (i.e. mean squared error). The connections between hidden units were constrained according to Dale's principle: neurotransmitters tend to be either excitatory or inhibitory such that the post-synaptic weights of each recurrent unit are all of the same sign[68]. 80% of units were fixed to be excitatory and the remaining 20% of units were inhibitory. The strong inhibitory signaling in the recurrent neural network enables stable temporal dynamics[67]. The cue contrast was set as 0.30 standard deviation (SD) and we trained RNNs, one at a time, with various target stimulus levels from zero to 0.13 SD, which mimics the near-threshold targets setting in the patient task. The RNNs were trained with 150,000 iterations and the model accuracy was tested on 50 batches of 50 sample trials for each target contrast, respectively. Finally, we fitted a psychometric detection curve by logistic function in the detection task and in the discrimination task, respectively. To verify the validity of the model, we compared the performance of trained models with an untrained model that was initialized with

random Gaussian connectivity weights without feeding stimuli inputs for learning. We also tested RNNs with different hidden units' size (128, 50, 32 and 16) and found out that those RNNs could not achieve human-level performance when the units were less than 16. To keep sufficient units for further clustering analysis, we adopted the number of hidden units N as 50 to balance the model's complexity and variability.

Trajectory k-means clustering on RNN unit activities

We applied the above-mentioned trajectory clustering method to classify dynamic patterns of the fifty hidden units for intermediate target contrast (0.10 SD). We generated 20 batches of 50 trials (1000 trials in total) and averaged unit activities by conditions (all conditions: validly cued seen, invalidly cued seen and no target). Unit trajectories were iteratively partitioned into 2-10 clusters across the three-dimensional condition space. The highest averaged silhouette value was obtained while the number of cluster equals to five (Fig S4d). We compared seen/unseen trials by time-resolved *t*-test in the post-target period (target onset to 1,000ms, see Fig 5d). Holm-Bonferroni correction was applied over all the time points. We also conducted the same clustering analysis on a reduced target contrast (0.03 and 0.06 SD). However, only a late cluster showed a significant consciousness effect. It might be due to the faint target contrast that other clusters did not attend to significance level.

Computing similarity between neural clusters and RNN clusters

In the four neural clusters showing the cue modulation on conscious perception (Visual, Sustained, Late accumulation and Reorienting), we averaged their neural activities in seen and in unseen trials. We then generated 500 batches of 50 trials with the RNN model simulation for both trained and untrained models. In each batch, we averaged the clusters' unit activity in seen and in unseen trials. The similarity was quantified by Pearson correlation coefficient between the

RNN clusters' temporal trajectory to the neural cluster ones, with the averaged coefficient of conditions. Therefore, we obtained a distribution of correlation coefficient with 500 samples for the trained and for the untrained models. Finally, we conducted a one-sided permutation test with 1,000 permutations to compare the two distributions, in order to identify the RNN clusters being significantly more similar to the neural clusters in the trained model than in the untrained model.

Directed connection weight graph

The intracerebral contacts of neural clusters were pooled from different patients, which limits the way to analyze their inter-cluster connections. However, the modeling provides the possibility to examine how unit clusters are inter-connected as an integrated model. To this end, we extracted unit input weights, directed unit-to-unit connection weights and output weights of the trained model (Fig. S4e). In each cluster, we observed both excitatory (E) and inhibitory (I) type units, of which the number was fully task-optimized without any prior. The Sustained cluster contained 16 E and 4 I units. The Late accumulation cluster had 6 E and 2 I units. The Reorienting cluster had 8 E and 3 I units. We grouped units of the same E/I type in each cluster, resulting in six groups in total. We then computed a directed cluster connection graph by averaging unit connection weights from one group ("pre-synaptic") to each of the other groups ("post-synaptic"). For example, to compute the connection weight from the Sustained excitatory group (16 units) to the Reorienting excitatory group (8 units), we averaged a total 128 directed unit-to-unit connection weights. One-sample *t*-test was performed with significance level controlled at $p < 0.5$, Holm-Sidak corrected for multi-comparison among groups of units.

Lesion analysis

To confirm the functional roles of the clusters, we lesioned units, one cluster at a time, and monitored the decrease in task performance. This was achieved by setting unit's connection

weights with inputs, all recurrent units and output to zero. We tested the task performance of the lesioned models with generated 50 batches of 50 trials.

Acknowledgements

This work is supported by the Agence Nationale de la Recherche through ANR-16-CE37-0005 and ANR-10-IAIHU-06, by the Fondation pour la Recherche sur les AVC through FR-AVC-017, by ICM-OCIRP, and by funding from Dassault Systèmes.

Author contributions

Conceptualization: DJB, PB

Data curation: KL

Formal analysis: JL, DJB

Methodology: JL, JDS, DJB, PB

Visualization: JL, JDS, TSM

Funding acquisition: PB, VN, CA

Project administration: PB

Resources: KL, SFV, VN, CA, VL

Software: JL, DJB, JDS, SFV

Supervision: TSM, PB

Writing - original draft: JL

Writing - review & editing: All authors

Competing interests:

Authors declare that they have no competing interests.

Materials & Correspondence

Requests for materials and correspondence should be addressed to Jianghao Liu and Paolo Bartolomeo.

References

1 Melloni, L., Mudrik, L., Pitts, M. & Koch, C. Making the hard problem of consciousness easier. *Science* **372**, 911-912 (2021).

2 Lamme, V. A. F. Challenges for theories of consciousness: seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **373** (2018).

3 Tononi, G., Boly, M., Massimini, M. & Koch, C. Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neurosci* **17**, 450-461 (2016).

4 Koch, C., Massimini, M., Boly, M. & Tononi, G. Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience* **17**, 307-321 (2016).

5 Sergent, C. *et al.* Bifurcation in brain dynamics reveals a signature of conscious processing independent of report. *Nature Communications* **12**, 1149 (2021).

6 Brown, R., Lau, H. & LeDoux, J. E. Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences* **23**, 754-768 (2019).

7 Mashour, G. A., Roelfsema, P., Changeux, J. P. & Dehaene, S. Conscious processing and the global neuronal workspace hypothesis. *Neuron* **105**, 776-798 (2020).

8 Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J. & Sergent, C. Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences* **10**, 204-211 (2006).

9 Buschman, T. J. & Kastner, S. From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron* **88**, 127-144 (2015).

10 Chun, M. M., Golomb, J. D. & Turk-Browne, N. B. A taxonomy of external and internal attention. *Annu Rev Psychol* **62**, 73-101 (2011).

11 Buschman, T. J. & Miller, E. K. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *science* **315**, 1860-1862 (2007).

12 Chica, A. B., Bartolomeo, P. & Lupiáñez, J. Two cognitive and neural systems for endogenous and exogenous spatial attention. *Behavioural brain research* **237**, 107-123 (2013).

13 Koch, C. & Tsuchiya, N. Attention and consciousness: two distinct brain processes. *Trends Cogn Sci* **11**, 16-22 (2007).

14 Maier, A. & Tsuchiya, N. Growing evidence for separate neural mechanisms for attention and consciousness. *Attention, perception & psychophysics* **83**, 558-576 (2021).

15 Tallon-Baudry, C. On the neural mechanisms subserving consciousness and attention. *Front Psychol* **2**, 397 (2011).

16 Chica, A. & Bartolomeo, P. Attentional Routes to Conscious Perception. *Frontiers in Psychology* **3** (2012).

17 Sergent, C. *et al.* Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Curr Biol* **23**, 150-155 (2013).

18 Chica, A. B., Lasaponara, S., Lupianez, J., Doricchi, F. & Bartolomeo, P. Exogenous attention can capture perceptual consciousness: ERP and behavioural evidence. *Neuroimage* **51**, 1205-1212 (2010).

19 Chica, A. B. *et al.* Spatial attention and conscious perception: the role of endogenous and exogenous orienting. *Attention, perception & psychophysics* **73**, 1065-1081 (2011).

20 Spagna, A. *et al.* The cost of attentional reorienting on conscious visual perception: an MEG study. *bioRxiv*, 2020.2012.2005.413161 (2022).

21 Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* **3**, 201-215 (2002).
22 Chica, A. B., Paz-Alonso, P. M., Valero-Cabré, A. & Bartolomeo, P. Neural Bases of the Interactions between Spatial Attention and Conscious Perception. *Cerebral Cortex* **23**, 1269-1279 (2012).
23 Bartolomeo, P. Visual neglect. *Current opinion in neurology* **20**, 381-386 (2007).
24 Bartolomeo, P., Thiebaut De Schotten, M. & Chica, A. B. Brain networks of visuospatial attention and their disruption in visual neglect. *Frontiers in Human Neuroscience* **6** (2012).
25 Corbetta, M., Patel, G. & Shulman, G. L. The reorienting system of the human brain: from environment to theory of mind. *Neuron* **58**, 306-324 (2008).
26 Thiebaut de Schotten, M. *et al.* A lateralized brain network for visuospatial attention. *Nat Neurosci* **14**, 1245-1246 (2011).
27 Bartolomeo, P. & Seidel Malkinson, T. Hemispheric lateralization of attention processes in the human brain. *Curr Opin Psychol* **29**, 90-96 (2019).
28 Chica, A. B., Martín-Arévalo, E., Botta, F. & Lupiáñez, J. The Spatial Orienting paradigm: how to design and interpret spatial attention experiments. *Neurosci Biobehav Rev* **40**, 35-51 (2014).
29 Green, D. M. & Swets, J. A. *Signal detection theory and psychophysics*. Vol. 1 (Wiley New York, 1966).
30 Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E. & Crone, N. E. High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. *Progress in neurobiology* **98**, 279-301 (2012).
31 Manning, J. R., Jacobs, J., Fried, I. & Kahana, M. J. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J Neurosci* **29**, 13613-13620 (2009).
32 Malkinson, T. S. *et al.* From perception to action: Intracortical recordings reveal cortical gradients of human exogenous attention. *bioRxiv*, 2021.2001.2002.425103 (2022).
33 Tzagarakis, C., Ince, N. F., Leuthold, A. C. & Pellizzer, G. Beta-band activity during motor planning reflects response uncertainty. *Journal of Neuroscience* **30**, 11270-11277 (2010).
34 Vu, A. T. *et al.* High resolution whole brain diffusion imaging at 7 T for the Human Connectome Project. *Neuroimage* **122**, 318-331 (2015).
35 Rojkova, K. *et al.* Atlasing the frontal lobe connections and their variability due to age and education: a spherical deconvolution tractography study. *Brain Structure and Function* **221**, 1751-1766 (2016).
36 Foley, J. M. & Legge, G. E. Contrast detection and near-threshold discrimination in human vision. *Vision research* **21**, 1041-1053 (1981).
37 Wutte, M. G., Smith, M. T., Flanagin, V. L. & Wolbers, T. Physiological signal variability in hMT+ reflects performance on a direction discrimination task. *Frontiers in psychology* **2**, 185 (2011).
38 Uka, T. & DeAngelis, G. C. Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *Journal of Neuroscience* **23**, 3515-3530 (2003).
39 Posner, M. Attention: The mechanism of consciousness. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 7398-7403 (1994).

40 Botta, F., Ródenas, E. & Chica, A. B. Target bottom-up strength determines the extent of attentional modulations on conscious perception. *Experimental brain research* **235**, 2109-2124 (2017).
41 Vugt, B. v. *et al.* The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* **360**, 537-542 (2018).
42 Vergani, F., Ghimire, P., Rajashekar, D., Dell'acqua, F. & Lavrador, J. P. Superior longitudinal fasciculus (SLF) I and II: an anatomical and functional review. *Journal of Neurosurgical Sciences* (2021).
43 Rozier, C. *et al.* Conscious and unconscious expectancy effects: A behavioral, scalp and intracranial electroencephalography study. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* **131**, 385-400 (2020).
44 Bartolomeo, P., Zieren, N., Vohn, R., Dubois, B. & Sturm, W. Neural correlates of primary and reflective consciousness of spatial orienting. *Neuropsychologia* **46**, 348-361 (2008).
45 Tecce, J. J. Contingent negative variation (CNV) and psychological processes in man. *Psychological bulletin* **77**, 73-108 (1972).
46 Thiebaut de Schotten, M. *et al.* Direct evidence for a parietal-frontal pathway subserving spatial awareness in humans. *Science* **309**, 2226-2228 (2005).
47 Thiebaut de Schotten, M. *et al.* Damage to white matter pathways in subacute and chronic spatial neglect: a group study and 2 single-case studies with complete virtual "in vivo" tractography dissection. *Cereb Cortex* **24**, 691-706 (2014).
48 Heekeren, H. R., Marrett, S. & Ungerleider, L. G. The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience* **9**, 467-479 (2008).
49 Summerfield, C. *et al.* Predictive codes for forthcoming perception in the frontal cortex. *Science* **314**, 1311-1314 (2006).
50 Kastner, S. & Ungerleider, L. G. The neural basis of biased competition in human visual cortex. *Neuropsychologia* **39**, 1263-1276 (2001).
51 Aron, A. R., Robbins, T. W. & Poldrack, R. A. Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences* **18**, 177-185 (2014).
52 Gigliotta, O., Seidel Malkinson, T., Miglino, O. & Bartolomeo, P. Pseudoneglect in Visual Search: Behavioral Evidence and Connectional Constraints in Simulated Neural Circuitry. *eNeuro* **4** (2017).
53 Joglekar, M. R., Mejias, J. F., Yang, G. R. & Wang, X.-J. Inter-areal balanced amplification enhances signal propagation in a large-scale circuit model of the primate cortex. *Neuron* **98**, 222-234 (2018).
54 Dehaene, S. & Changeux, J.-P. Ongoing Spontaneous Activity Controls Access to Consciousness: A Neuronal Model for Inattentional Blindness. *PLOS Biology* **3**, e141 (2005).
55 Colás, I. *et al.* Conscious perception in patients with prefrontal damage. *Neuropsychologia* **129**, 284-293 (2019).
56 Pollack, I. & Norman, D. A. A non-parametric analysis of recognition experiments. *Psychonomic Science* **1**, 125-126 (1964).
57 Pérez-García, F. *et al.* Automatic segmentation of depth electrodes implanted in epileptic patients: a modular tool adaptable to multicentric protocols. *Epilepsia* **56**, 227 (2015).

58 Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968-980 (2006).
59 Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience* **2011**, 156869 (2011).
60 Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **9** (2008).
61 Wang, R., Benner, T., Sorensen, A. G. & Wedeen, V. J. in *Proc Intl Soc Mag Reson Med.* (Berlin).
62 Foulon, C. *et al.* Advanced lesion symptom mapping analyses and implementation as BCBtoolkit. *GigaScience* **7** (2018).
63 Barak, O. Recurrent neural networks as versatile tools of neuroscience research. *Current opinion in neurobiology* **46**, 1-6 (2017).
64 Kriegeskorte, N. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science* **1**, 417-446 (2015).
65 Yamins, D. L. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci U S A* **111**, 8619-8624 (2014).
66 Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T. & Wang, X. J. Task representations in neural networks trained to perform many cognitive tasks. *Nat Neurosci* **22**, 297-306 (2019).
67 Kim, R. & Sejnowski, T. J. Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks. *Nat Neurosci* **24**, 129-139 (2021).
68 Song, H. F., Yang, G. R. & Wang, X.-J. Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework. *PLOS Computational Biology* **12**, e1004792 (2016).
69 Ehrlich, D. B., Stone, J. T., Brandfonbrener, D., Atanasov, A. & Murray, J. D. PsychRNN: An Accessible and Flexible Python Package for Training Recurrent Neural Network Models on Cognitive Tasks. *Eneuro* **8** (2021).