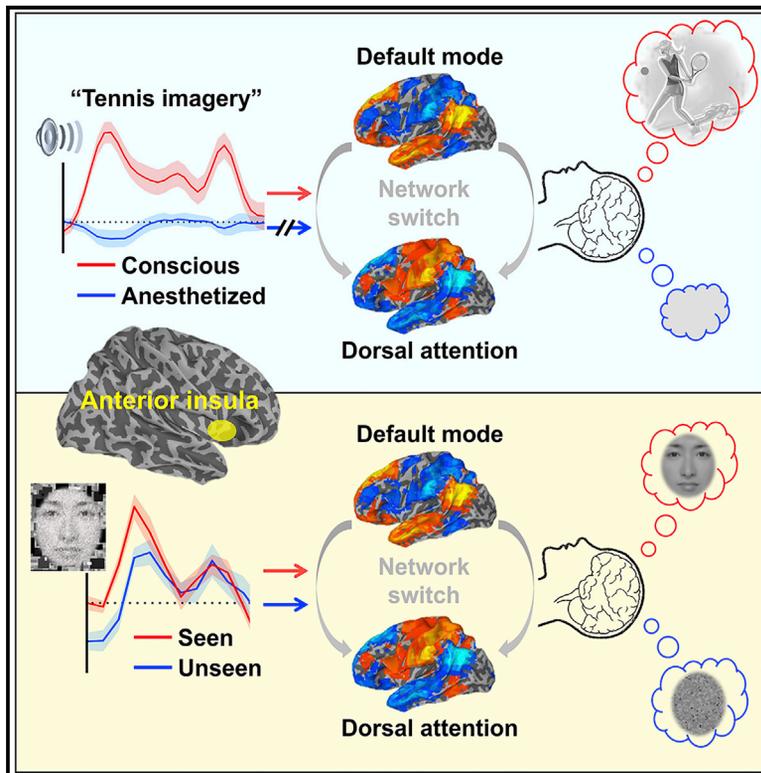# Cell Reports

Anterior insula regulates brain network transitions that gate conscious access

Graphical abstract

Authors

Zirui Huang, Vijay Tarnal,
Phillip E. Vlisides, ..., Paul Picton,
George A. Mashour, Anthony G. Hudetz

Correspondence

huangzu@umich.edu (Z.H.),
ahudetz@med.umich.edu (A.G.H.)

In brief

In a human neuroimaging study, Huang
et al. manipulate the level and content of
consciousness using independent
experimental protocols to demonstrate
that the anterior insula, situated between
unimodal and transmodal cortical areas
along the brain's functional hierarchy,
serves as a gate for conscious access of
sensory information.

Highlights

- Dysfunction of anterior insula during anesthesia disables brain network transitions

- Prestimulus activity of anterior insula predicts conscious access of visual stimuli

- Anterior insula might be a gate for conscious access of sensory information

- This cortical gate occupies an intermediate position along a neurocognitive hierarchy

CellPress

Cell Reports

Article

Anterior insula regulates brain network transitions that gate conscious access

Zirui Huang,[1,2,*] Vijay Tarnal,[1,2] Phillip E. Vlisides,[1,2] Ellen L. Janke,[1,2] Amy M. McKinney,[1] Paul Picton,[1] George A. Mashour,[1,2,3] and Anthony G. Hudetz[1,2,3,4,*]
[1]Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI 48109, USA
[2]Center for Consciousness Science, University of Michigan Medical School, Ann Arbor, MI 48109, USA
[3]Neuroscience Graduate Program, University of Michigan, Ann Arbor, MI 48109, USA
[4]Lead contact
*Correspondence: huangzu@umich.edu (Z.H.), ahudetz@med.umich.edu (A.G.H.)
https://doi.org/10.1016/j.celrep.2021.109081

SUMMARY

Conscious access to sensory information is likely gated at an intermediate site between primary sensory and transmodal association cortices, but the structure responsible remains unknown. We perform functional neuroimaging to determine the neural correlates of conscious access using a volitional mental imagery task, a report paradigm not confounded by motor behavior. Titrating propofol to loss of behavioral responsiveness in healthy volunteers creates dysfunction of the anterior insular cortex (AIC) in association with an impairment of dynamic transitions of default-mode and dorsal attention networks. Candidate subcortical regions mediating sensory gating or arousal (thalamus, basal forebrain) fail to show this association. The gating role of the AIC is consistent with findings in awake participants, whose conscious access is predicted by pre-stimulus AIC activity near perceptual threshold. These data support the hypothesis that AIC, situated at an intermediate position of the cortical hierarchy, regulates brain network transitions that gate conscious access.

INTRODUCTION

Despite decades of research in psychology and neuroscience, the question of why certain sensory stimuli are consciously perceived while others are not remains elusive. Empirical data suggest that conscious representations are constructed in the cortex (Dehaene and Changeux, 2011; Mashour and Hudetz, 2018; Mashour et al., 2020a) and that the entry of sensory inputs to the cortex is controlled or gated by the thalamus (Alkire et al., 2008; Brown et al., 2011; Halassa and Kastner, 2017; Liu et al., 2013; Mashour and Alkire, 2013; Sherman, 2016; Suzuki and Larkum, 2020). However, in unconscious subjects, sensory thalamocortical transmission may be preserved, yet stimuli received in the cortex are not perceived, as evidenced by a failure to adequately report or respond, which suggests a breakdown of conscious access (Boveroux et al., 2010; Davis et al., 2007; Gross et al., 2019; Hudetz, 2006; Mashour et al., 2020a). Here, we refer to conscious access as the global availability of sensory information to cognitive processors (Mashour et al., 2020a) such as those that mediate functions like working memory, verbal report, or motor behavior. Thus, conscious access occurs beyond the stage of primary sensory processing. We hypothesized that there exists a critical structure between primary sensory and association cortices where conscious access to sensory information is gated.

A candidate brain area may be situated at an intermediate position along the brain's functional hierarchy, where unimodal and transmodal operations interface. The anterior insular cortex (AIC) has been recognized as a central informational hub of the brain because it receives inputs from different sensory modalities and the internal environment (e.g., interoception or emotions), and it determines the relevance and processing priorities across modalities (Craig, 2009; Menon, 2011; Michel, 2017; Sterzer and Kleinschmidt, 2010; Uddin, 2015). Hence, it is plausible to hypothesize that the AIC may play a gating role for transmodal integration of information associated with conscious access. However, whether the AIC is the primary gate that controls conscious access has not been decisively determined using a rigorous experimental paradigm. Answering this question is arguably critical to understand the neural mechanisms underlying consciousness.

One challenge to making inferences about conscious access in typical study paradigms is the need for a motor response. This can confound investigations of the neural correlates of consciousness (Aru et al., 2012). Volitional mental imagery is now a well-established paradigm that can assess conscious representations in the absence of motor behavior; the resultant imagery effort is a surrogate of response reflecting conscious access. While this approach has primarily been employed in the detection of covert consciousness in patients with neuropathology (Monti et al., 2010; Owen et al., 2006), we have reproduced this finding in volunteers undergoing sedation (Huang et al., 2018a), enabling a behavior-free approach to identify a potential gate of conscious processing—from primary data (e.g., acoustic response to verbal instructions) to conscious access (e.g., content of imagery)—in the healthy brain.
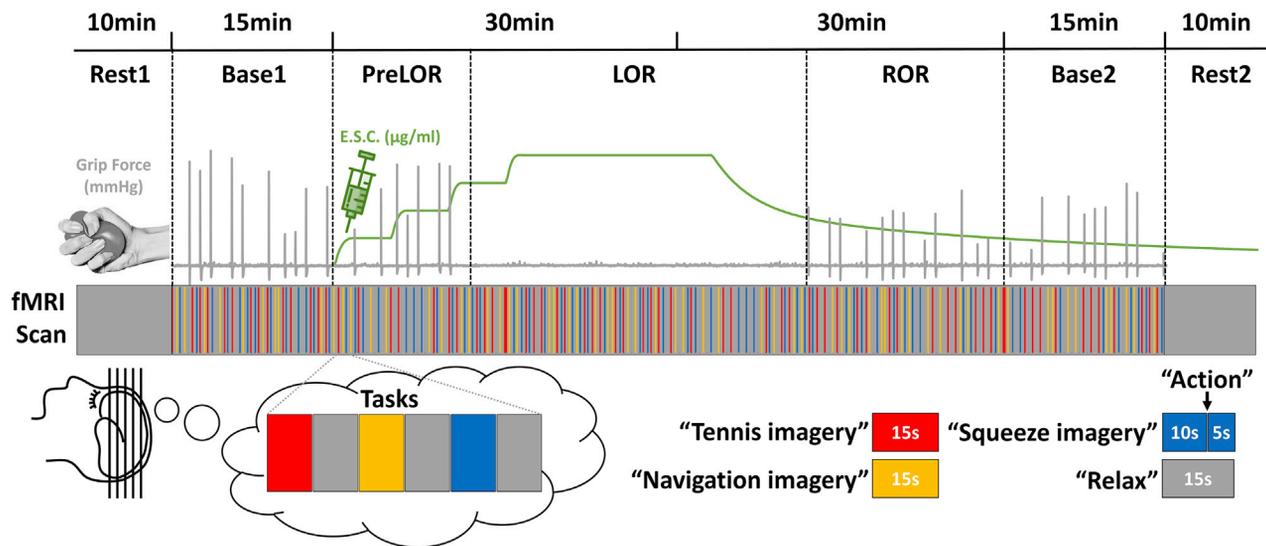
Figure 1. Experimental design

Schematic of the experimental protocol for stepwise intravenous infusion of propofol and fMRI tasks. The infusion rate was adjusted to achieve stepwise increasing target effect-site concentrations (ESCs) in 0.4-μg/ml increments. The final target concentration was one increment above that which first resulted in loss of behavioral responsiveness. The final target was then maintained at this level for approximately 22 min. After that, the infusion was terminated to allow for spontaneous recovery. Mental imagery and motor response tasks were tested before, during, and after propofol infusion. Subjects were asked to perform three imagery tasks (tennis, navigation, and hand squeeze) plus a motor response task (actual hand squeeze). The timing of "action" instructions and the actual motor response were used to determine the periods during which a participant retained responsiveness (PreLOR), loss of responsiveness (LOR), and recovery of responsiveness (ROR). Two 10-min resting-state baseline and two 15-min task baseline recordings were done before (Rest1 and Base1) and after (Base2 and Rest2) propofol infusion.

In this study, we performed functional magnetic resonance imaging (fMRI) in two independent experimental paradigms to identify the cortical gate for conscious access of sensory information. In the first experiment, we applied the general anesthetic propofol as a tool to modulate the level of consciousness of healthy volunteers instructed to engage in mental imagery tasks. We aimed to determine which brain region(s) played a key role in disabling the dynamic brain network transitions that have been associated with conscious processing (Huang et al., 2020). In the second experiment, we evaluated the identified critical brain area in a psychological setting where near-threshold perceptual awareness was studied in a classical backward-masking paradigm in awake healthy participants. Here, we aimed to determine whether the spontaneous activity of the presumed cortical gate was predictive of future conscious access (i.e., if the stimulus crossed or failed to cross the perceptual threshold). We report that both sets of experiments support a key role of the AIC in gating conscious access.

RESULTS

Pharmacological interventions to probe conscious access

Twenty-six healthy volunteers were studied using fMRI during and after intravenous propofol infusion (Figure 1). Participants were asked to perform three mental imagery tasks (tennis, navigation, and hand squeeze) plus a motor response task (squeezing a rubber ball by hand). A pseudo-randomized block design was applied

comprising 15 s of tennis imagery, 15 s of navigation imagery, and 10 s of squeeze imagery, followed by hand squeeze within 5 s after hearing the instruction and alternated with 15 s of rest. The propofol infusion rate was adjusted to achieve stepwise increasing target effect-site concentrations (ESCs) in 0.4-μg/ml increments. The final target concentration was one increment above that which first resulted in loss of behavioral responsiveness. The final target was then maintained at this level for approximately 22 min (see Method details for variations of the protocol). The infusion was then terminated to allow for spontaneous recovery. Behavioral responsiveness defined the periods during which a participant retained responsiveness (PreLOR), loss of responsiveness (LOR), and recovery of responsiveness (ROR). Two 10-min resting-state baseline and two 15-min task baseline recordings were done before (Rest1 and Base1) and after (Base2 and Rest2) propofol infusion.

Whole-brain task activations

Mental imagery and motor response tasks produced distinguishable patterns of activation in specific regions of the brain when participants were fully conscious before propofol infusion (during Base1). Consistent with previous reports (Huang et al., 2018a; Monti et al., 2010; Owen et al., 2006), positive activations (hereafter referred to as "activations") included the supplementary motor area (SMA) and precuneus (PreCu) during tennis imagery, the parahippocampal place area (PPA) and PreCu during navigation imagery, the SMA during squeeze imagery, and the primary motor cortex (M1) and SMA during motor response (Figure 2).
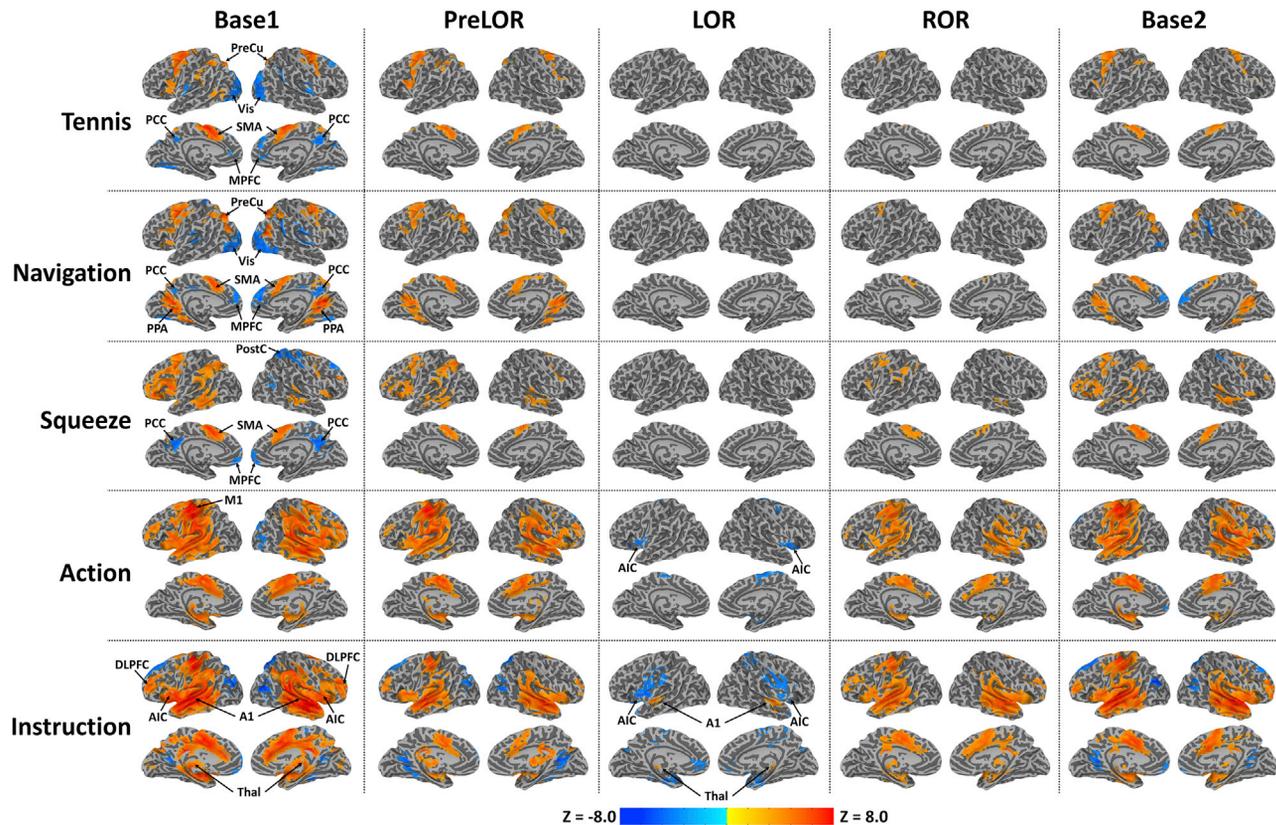
Figure 2. Task-induced brain activity

Group-level z-maps are shown for tennis imagery (Tennis; 15-s duration), navigation imagery (Navigation; 15-s duration), squeeze imagery (Squeeze; 10-s duration), actual hand squeeze (Action; 2-s duration), and verbal instruction (Instruction; 2-s duration) during Base1 (n = 26), PreLOR (n = 26), LOR (n = 26), ROR (n = 16), and Base2 (n = 25). All z-maps (one sample t test against zero) were corrected at the cluster level $\alpha < 0.05$. During LOR, mental-imagery-related activations were absent. Verbal-instruction-evoked activations were attenuated and constrained within the Thal and A1, and deactivations were seen in bilateral AIC. SMA, supplementary motor area; PreCu, precuneus; PPA, parahippocampal place area; M1, primary motor cortex; MPFC, medial prefrontal cortex; PCC, posterior cingulate cortex; Vis, visual cortex; PostC, postcentral gyrus; Thal, thalamus; A1, auditory cortex; AIC, anterior insular cortex; DLPFC, dorsal lateral PFC.

Negative activations (hereafter referred to as "deactivations") were found in the medial prefrontal cortex (MPFC) and posterior cingulate cortex (PCC) during all three imagery tasks. In addition, deactivation in the lateral occipital cortex (Vis) was seen during tennis and navigation imagery, and deactivation in the postcentral gyrus (PostC) was seen during squeeze imagery. Deactivations may indicate neural resource reallocations among competing systems (Anticevic et al., 2012; Raichle, 2015). Lastly, verbal instructions (i.e., "tennis imagery," "navigation imagery," "squeeze imagery," "action," and "relax") elicited widespread cortical and subcortical activations including the thalamus (Thal), primary auditory cortex (A1), AIC, and dorsal lateral PFC (DLPFC). Verbal instructions also elicited deactivations in the MPFC, PCC, PostC, and Vis. The aforementioned brain regions (defined by Base1) were used as regions of interest (ROIs) in the subsequent analyses (see Table S1 for a summary of ROIs).

During propofol infusion PreLOR, mental-imagery-related activations were mostly preserved, whereas deactivations were sparsely seen. During LOR, mental-imagery-related activations were absent, while verbal instruction-evoked activations were

attenuated and constrained within the Thal and A1. Instruction-evoked deactivations in the bilateral AIC were observed during LOR. After spontaneous emergence (ROR), activations evoked by motor response and verbal instruction were recovered, whereas mental-imagery-related activations were less prominent. During Base2, the patterns of activation in different tasks mostly mirrored those seen during Base1 and PreLOR.

Classifying the ROIs according to cortical gradients

We applied diffusion map embedding (Margulies et al., 2016) to characterize macroscale cortical gradients across different conditions. During the baseline condition (Base1), the first axis (gradient 1) depicts a gradient ranging from unimodal primary sensory areas (e.g., visual, auditory, somatosensory, and motor) to transmodal cortex (e.g., frontoparietal and default mode). The second axis (gradient 2) depicts a gradient running from the visual areas to somatomotor cortex (Figures 3A and 3B). These gradients are in agreement with those reported previously in the literature (Bethlehem et al., 2020; Hong et al., 2019; Margulies et al., 2016; Mckeown et al., 2020; Paquola et al., 2019; Vos de Wael
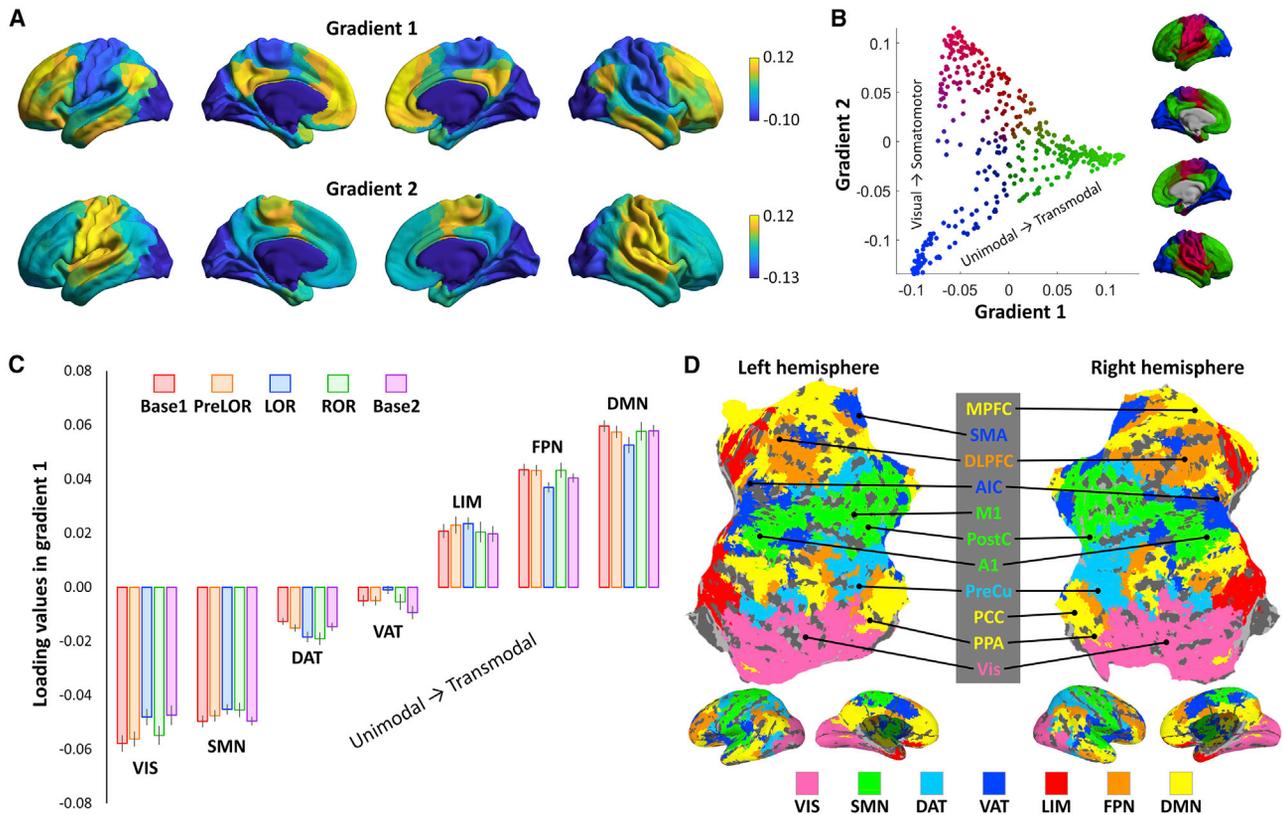
Figure 3. Cortical gradients of functional organization

(A) Topographic profiles of the first two gradients along the cortex during the baseline condition (Base1). See Figure S1 for other conditions.

(B) The two gradients are projected into a two-dimensional gradient space. The axes represent each gradient and separate distinct functional poles of cortical organization (i.e., unimodal to transmodal regions in gradient one and visual to somatomotor regions in gradient two).

(C) The distribution of gradient eigenvector loading values are shown for Base1 (n = 26), PreLOR (n = 26), LOR (n = 26), ROR (n = 16), and Base2 (n = 25) in seven pre-defined functional networks including the default-mode network (DMN), frontoparietal network (FPN), limbic network (LIM), ventral attention/salience network (VAT), dorsal attention network (DAT), somatomotor network (SMN), and visual network (VIS). Error bars indicate ± SEM across subjects.

(D) The locations (cluster peaks; Table S1) of ROIs during the task are marked within these networks.

et al., 2020). In addition, the cortical gradients were, overall, similar across different conditions (Figure S1). We further characterized the first principle cortical gradient (gradient 1) at the network level by plotting the distribution of gradient eigenvector loading values across seven pre-defined functional networks (Yeo et al., 2011). We then localized the ROIs within these networks—namely, the MPFC, PCC, and PPA situated in the default-mode network (DMN); the DLPFC situated in the frontoparietal network (FPN); the AIC and SMA situated in the ventral attention/salience network (VAT); the PreCu situated in the dorsal attention network (DAT); the M1, PostC, and A1 situated in the somatomotor network (SMN); and the lateral occipital cortex (Vis) situated in the visual network (VIS).

Task activations in the ROIs

To determine the processing stage along the cognitive hierarchy at which the task-induced brain activity breaks down during LOR, we extracted the time course of fMRI signal time locked to the onset of mental imagery tasks in the above-defined ROIs. In addition, we included a region in the basal forebrain containing cholinergic cells (Ch4) that are known to participate in

arousal regulation (Brown et al., 2011; Liu et al., 2018; Turchi et al., 2018). Based on available evidence for the putative functional role of specific brain regions in the hierarchy of cognitive processing (Brown et al., 2019; Menon, 2011), the ROIs were considered to encompass bottom-up sensory relays (Thal), arousal regulatory area (Ch4), cortical sensory (A1) and motor (M1) regions, key nodes of ventral attention/salience processing (AIC), high-order cognitive processors (PPA, SMA, and PreCu), and top-down control (DLPFC) (Figure 4).

Compared to that of conscious conditions (Base1, PreLOR, ROR, and Base2), instruction-evoked activation in the Thal and A1 was substantially reduced during LOR. Activation in Ch4 was clearly seen during motor response but not during LOR when overt motor response was absent. Note that the signal-to-noise ratio in Ch4 was low because this region is very small (Liu et al., 2018). Importantly, the deactivations in the AIC indicated a functional failure that occurred at an intermediate position between unimodal and transmodal areas along the cortical gradients. Accordingly, other brain regions presumably associated with higher-order cognitive processes such as the PPA, SMA, PreCu, and DLPFC were inactive during LOR.
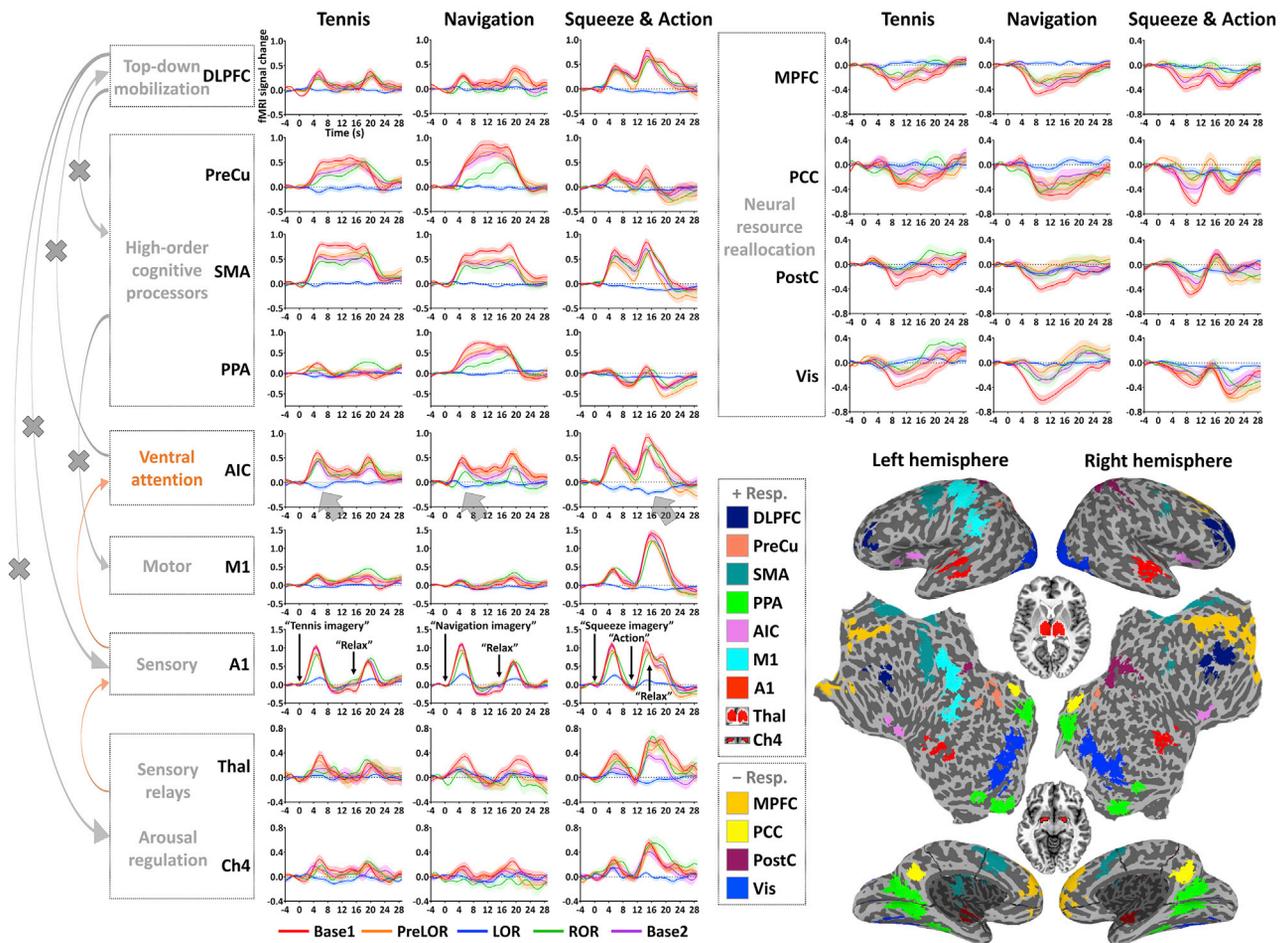
Figure 4. Time course of fMRI signal change in ROIs
Each time course includes 4.0-s pre-stimulus baseline and 29.6-s post-stimulus period (step: 0.8 s). The fMRI signal is corrected by subtracting the mean value of the pre-stimulus baseline. Time courses are plotted for Base1 (n = 26), PreLOR (n = 26), LOR (n = 26), ROR (n = 16), and Base2 (n = 25). Shaded areas indicate ± SEM across subjects. Brown arrows versus gray arrows on the left indicate preserved versus disrupted cognitive processing pathways during LOR. The deactivations in the AIC represent a functional failure at an intermediate position in the brain's functional hierarchy. The ROIs are mapped on the inflated and flattened cortical surface (except for Thal and Ch4; see small horizontal sections in the middle). The +Resp. and −Resp. indicate regions showing activation and deactivation, respectively.

AIC controls macroscale brain network switches

We sought to track time-locked macroscale brain network transitions required for conscious processing (i.e., switches between DMN and DAT) (Huang et al., 2020) and used this to identify the critical brain regions that mediate such transitions. This was achieved by quantifying the spatial similarity between the signal intensity of each fMRI volume and pre-defined centroids of co-activation patterns (CAPs) derived from our previous study (Huang et al., 2020). Specifically, the CAPs were identified by an unsupervised machine-learning approach (i.e., k-means clustering algorithm) that assessed areas across the brain that are consistently activated together rather than averaging activity over long periods (Huang et al., 2020). The CAPs were, overall, in agreement with the aforementioned functional networks (Yeo et al., 2011) except that the limbic network was not identified, and two other networks representing global brain activation

and deactivation (GN+ and GN−) were included. Of note, the CAPs should not be considered as conventionally defined brain network templates, as they do not isolate specific brain areas or networks. Instead, they consist of four pairs of whole-brain "mirror" motifs (i.e., an anti-phasic topography). For instance, the DMN+ (according to the definition of CAPs) is accompanied by co-deactivation of DAT (DAT−), and vice versa for DAT+ (DMN−). We used this approach for two reasons: (1) it tracks the anti-correlation relationship between DMN and DAT without the need of temporal averaging or applying global signal regression (GSR) (Huang et al., 2020); and (2) it quantifies whole-brain large-scale topography, so it is relatively insensitive to undesired signal variability in focal brain areas. Moreover, to ensure an unbiased estimation, we included not only the DMN+ and DAT+, but also other CAPs including the FPN+, SMN+, VIS+, VAT+, GN+, and GN−.
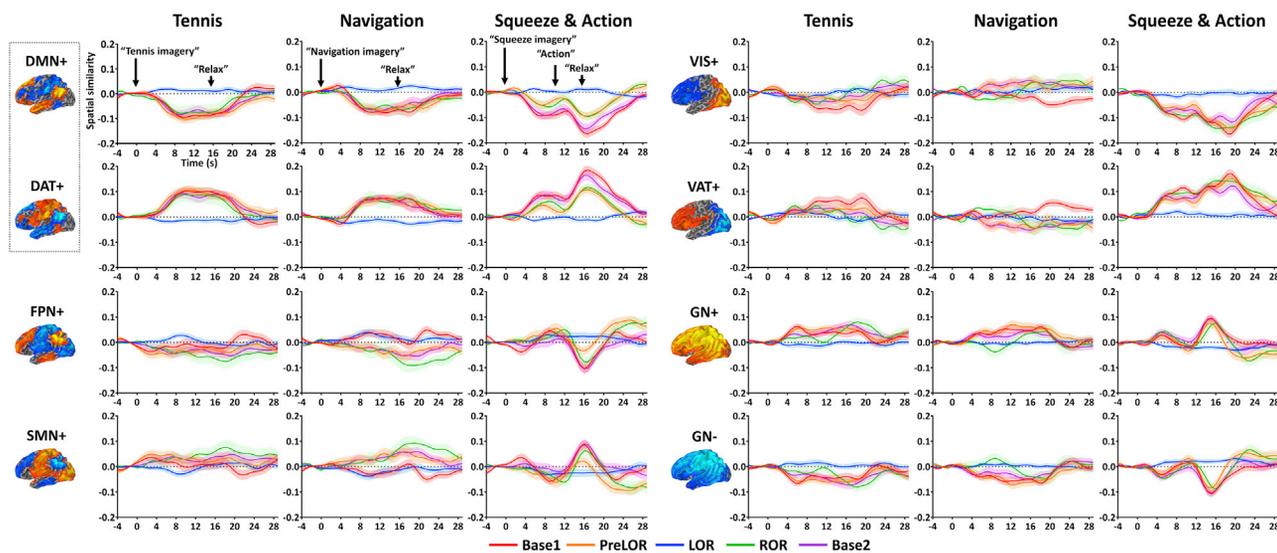
Figure 5. Time course of spatial similarity of eight CAPs
Each plot includes 4.0-s pre-stimulus baseline and 29.6-s post-stimulus period (step: 0.8 s). The spatial similarity values were corrected by subtracting the mean value of the pre-stimulus baseline. Time courses are plotted for Base1 (n = 26), PreLOR (n = 26), LOR (n = 26), ROR (n = 16), and Base2 (n = 25). Shaded areas indicate ± SEM across subjects. The CAPs include the DMN+, DAT+, FPN+, SMN+, VIS+, VAT+, and global network of activation and deactivation (GN+ and GN–). During conscious conditions (Base1, PreLOR, ROR, and Base2), there were positive (and negative) modulations in the DAT+ (and DMN+) for all imagery tasks. Other CAPs were less engaged in mental imagery tasks, expect for the VAT+ and VIS+ during squeeze imagery. The DMN-DAT switch was abolished during LOR.

For each CAP, the spatial similarity time courses were averaged across task blocks (Figure 5). As the fMRI-derived blood-oxygenation-level-dependent (BOLD) response relies on changes in the cerebral vasculature, it exhibits a temporal delay. In general, BOLD signal changes take 6–12 s to reach maximum intensity and can remain relatively constant for sustained periods of activity such as in an fMRI block design (15-s periods of mental imagery in our case). In addition, omitting a few seconds following the block onset and including a few seconds following the block offset have been recommended for frame-wise analysis (LaConte et al., 2007). Hence, we calculated the area under the curve (AUC) of 4 to 20 s following verbal instruction (corrected by pre-stimulus baseline period of −4.0 to −0.8 s) to quantify the overall temporal gain of CAP transitioning. The AUC is hereafter referred to as modulation index.

During conscious conditions (Base1, PreLOR, ROR, and Base2), there was significant positive modulation in the DAT+ (p < 0.001) and significant negative modulation in the DMN+ (p < 0.001) for all imagery tasks. This was consistent with the known DMN-DAT switch when participants engaged in a cognitive task (Raichle, 2015). Other CAPs appeared to be less engaged in mental imagery tasks, expect for the VAT+ and VIS+ during squeeze imagery. Moreover, confirming our predictions (Huang et al., 2020), the DMN-DAT switch was abolished during LOR.

We next tested our hypothesis that the AIC plays a role in mediating network transitions associated with conscious access. We performed inter-subject correlation analysis between instruction-evoked activation in the AIC and a joint measure—namely, DAT-DMN modulation index defined as DAT+ modulation index minus DMN+ modulation index (given their anti-phasic

relationship)—and averaged across the three imagery tasks. Supporting our hypothesis, instruction-evoked activation in the AIC positively correlated with the DAT-DMN modulation index (r = 0.58, p < 0.001; Figure 6A). That is, a higher AIC activity level following a brief verbal instruction was associated with a larger gain of DMN-DAT switch when a sustained mental imagery task was performed.

To examine how specific the role of AIC was (versus other brain regions) during the DMN-DAT switch, we performed voxel-wise correlation analysis between the instruction-evoked whole-brain activation and the DAT-DMN modulation index. Although we found positive correlations across widespread cortical and subcortical regions (Figure 6B), such correlations may have been confounded by sensory- and arousal-related processes due to the verbal instruction. Therefore, we performed a partial correlation analysis by including the activations of A1, Ch4, and Thal (extracted from pre-defined ROIs) as covariates. We found that the bilateral AICs were the only areas that positively (and with statistical significance) correlated with the DAT-DMN modulation index (Figure 6C). Taken together, these results suggest a primary role for the AIC in mediating the DMN-DAT switching necessary for conscious access (see Figure 6D for a schematic illustration).

To further support the AIC's role in mediating the DMN-DAT switch, we analyzed functional connectivity among the ROIs (Figure S2). For methodological considerations, the analysis was performed both without and with applying GSR procedure (non-GSR versus GSR), as the two approaches could reveal complementary views of the brain's functional organization (Murphy and Fox, 2017). For non-GSR, LOR was accompanied by a global reduction of functional connectivity comparing to
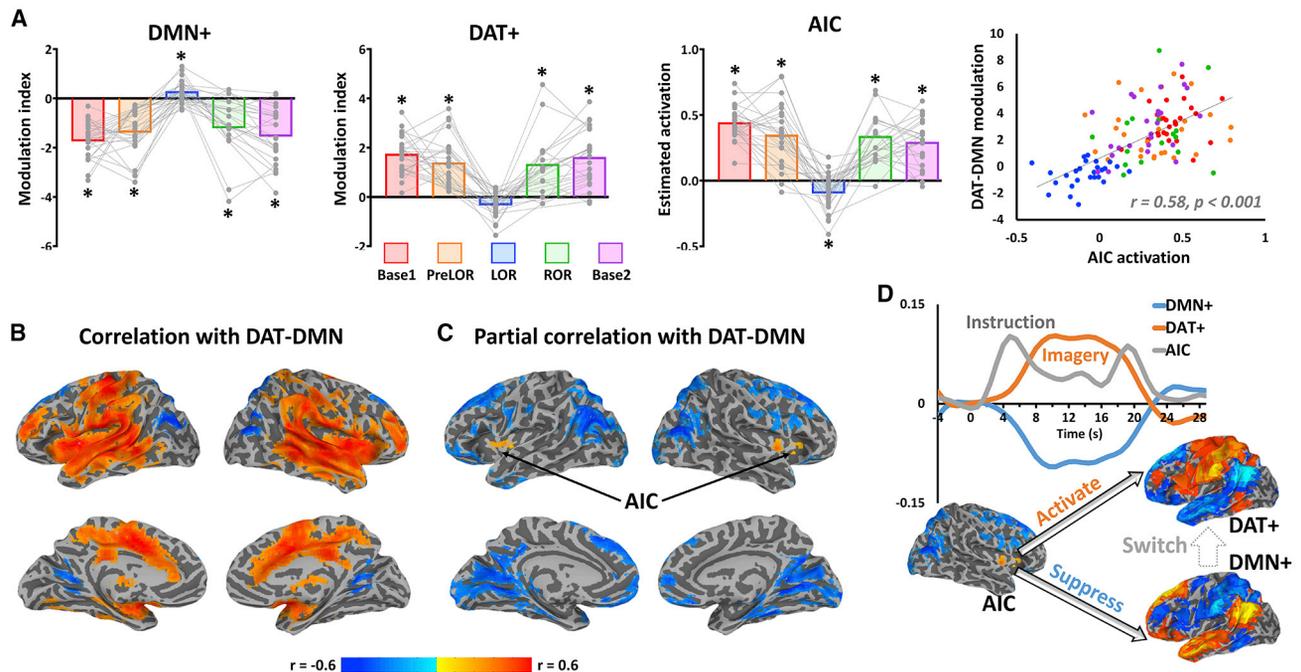
Figure 6. AIC controls the DMN-DAT switches

(A) Modulation indices were plotted for DMN+ and DAT+. Instruction-evoked activation estimated from general linear model was plotted for the AIC. Each gray dot represents an individual participant during Base1 (n = 26), PreLOR (n = 26), LOR (n = 26), ROR (n = 16), or Base2 (n = 25) connected by gray lines across conditions. Bars represent the group averages for each condition. Asterisk indicates statistical significance (one sample t test against zero) at false discovery rate (FDR)-corrected alpha < 0.05. Pearson correlation was performed between instruction-evoked activation in the AIC and DAT-DMN modulation index across subjects and conditions (n = 119).

(B) Voxel-wise correlation between whole-brain instruction-evoked activation and DAT-DMN modulation index.

(C) Voxel-wise partial correlation between whole-brain instruction-evoked activation and DAT-DMN modulation index by including the activations of A1, Ch4, and Thal as covariates.

(D) Schematic illustration of hypothesized conscious processing. The AIC initiates a large-scale network transition by activating the DAT and suppressing the DMN. The group-averaged time courses are shown as an example for the AIC's activity (arbitrary unit on y axis for illustrative purpose) and DMN+ and DAT+ spatial similarity during tennis imagery in the baseline condition.

conscious conditions (Rest1, Base1, PreLOR, ROR, Base2, and Rest2). This result was in line with our previous reports (Huang et al., 2018b; Tanabe et al., 2020). Using GSR, which removes the global component of signal correlations, we found diminished anti-correlation between the AIC and the core regions of the DMN (MPFC and PCC) during LOR. This provides evidence that the failure of the AIC was associated with a lack of inhibitory regulation of DMN.

Probing the causal role of the AIC in conscious access

We further evaluated the AIC's role in conscious access in a psychological setting. Participants (n = 19) were briefly shown either a face or a scrambled face image, followed by a high-contrast image (mask) using a classical backward-masking paradigm. For example, in a near-threshold condition, both face and scrambled face were presented very rapidly (e.g., 33 ms) and immediately replaced by a mask for a longer duration (e.g., 400 ms), such that the mask could interrupt conscious processing of the initial stimulus. We determined the supraliminal (above threshold) and near-threshold stimulus presentations by manipulating the duration of the target stimuli. For the supraliminal condition, a 200-ms target duration was used. Individual thresholds

for discriminating a face from a scrambled face were determined by an adaptive staircase procedure. The participants were asked whether they recognized a face or not during both near-threshold and supraliminal conditions (Figures 7A–7D).

Because both pre-stimulus (indexing spontaneous variations) and post-stimulus (indexing access consciousness) brain activity were of interest in this experiment, conventional general linear model analysis was not performed. Using a whole-brain analysis, we first contrasted the post-stimulus activity of seen versus unseen of a face in the near-threshold condition. Unlike the fMRI block design in which the BOLD response is sustained, we chose a sparse event-related design with a much shorter BOLD response duration. Accordingly, we calculated the response AUC from the 1.5–6.0-s period following stimulus presentation. This time window was chosen to include the rise-apex period of BOLD response while excluding the stimulus onset (t = 0). We found that higher post-stimulus activity in the AIC, as well as the DLPFC, anterior cingulate cortex (ACC), and Thal, was associated with conscious access of seeing a face (Figures 7E and 7F; see Figure S3 for supraliminal conditions). The structural boundaries of the AIC identified in this experiment coincided well with the structural boundaries of the AIC identified in participants
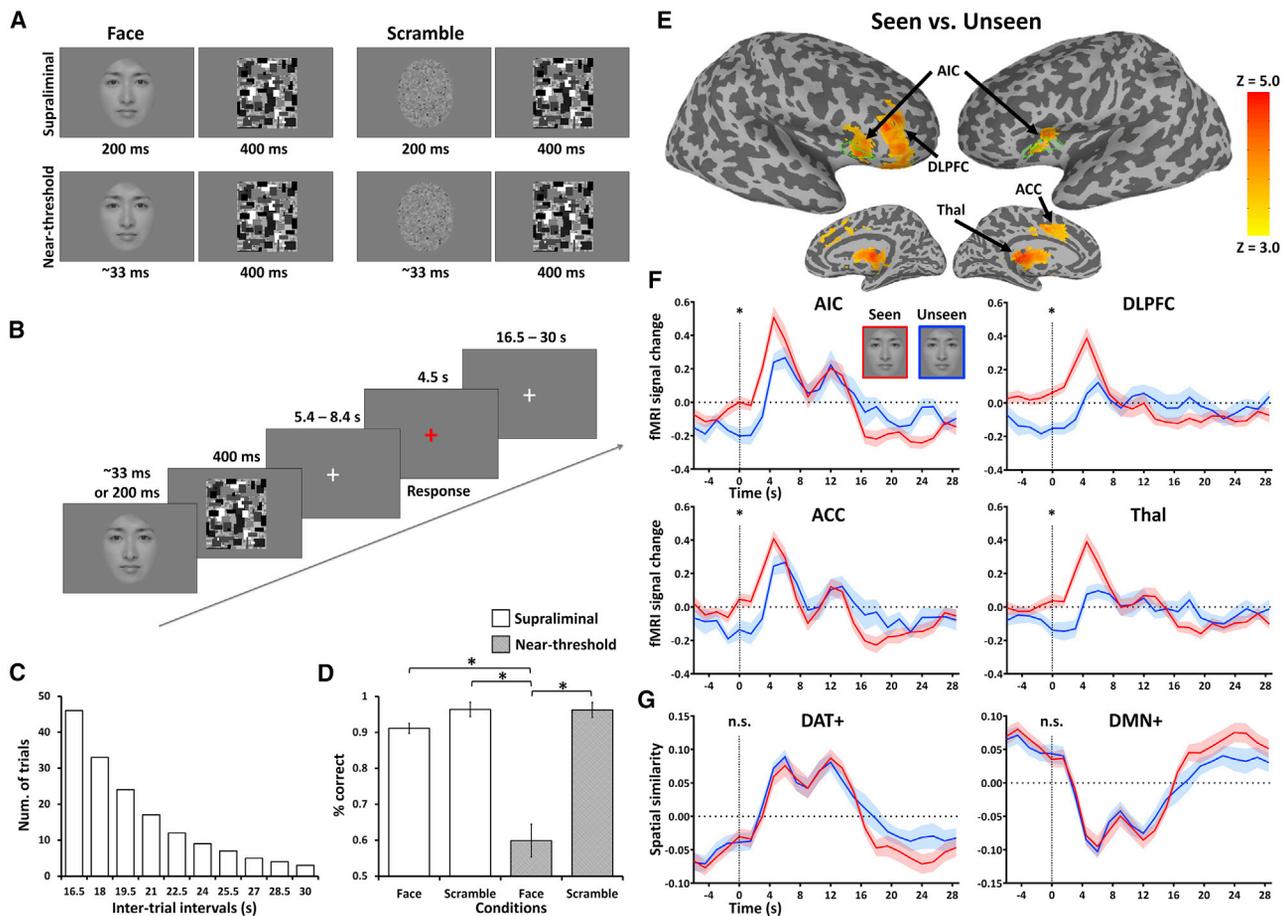
Figure 7. Testing conscious access in a psychological setting

(A) A face or a scrambled face was briefly displayed and then masked with a high-contrast image. Display duration of 200 ms was used for supraliminal presentation. The near-threshold face presentation was individualized by an adaptive staircase procedure. The threshold duration was 33 ms in 17 out of 19 participants (see Method details).

(B) Each trial started with a brief flash of face or scrambled face image. Participants were instructed to view the stimuli but not respond until a red fixation cross prompt appeared on the screen. They were required to report whether they had seen a face or not. After their button press response, an unpredictably long rest period with a white fixation cross was used to separate trials (19.5-s mean duration; 1.5-s steps).

(C) The duration of the rest periods followed an exponential distribution.

(D) Behavioral results. The hit rates (p[present|present]) and correct rejections (p[absent|absent]) of a face were 91% (SD = 5.9%) and 96% (SD = 8.5%) in supraliminal conditions and 60% (SD = 19.3%) and 96% (SD = 8.8%) in near-threshold conditions. Significant differences in those rates were found between the near-threshold face and all other conditions (*p < 0.001).

(E) Group-level z-maps of stimulus-induced activity for near-threshold seen versus unseen of a face. The z-maps were thresholded at cluster level $\alpha < 0.05$.

(F) Time courses of fMRI signal change for near-threshold seen versus unseen in the AIC, DLPFC, ACC, and Thal.

(G) Time courses of spatial similarity for DAT+ and DMN+. *p < 0.01 (paired sample t tests); n.s., non-significance. Shaded areas and error bars indicate ± SEM across subjects (n = 19).

receiving propofol in the first experiment (delineated by green edges in Figure 7E).

Importantly, this perceptual bias also occurred at the stimulus onset (t = 0) for all these regions (AIC, p < 0.001; DLPFC, p < 0.001; ACC, p = 0.006; Thal, p = 0.005; paired sample t tests). These effects at the stimulus onset could only be explained by pre-stimulus (rather than post-stimulus) activity considering the hemodynamic response delay of fMRI signal, thus indicating a causal impact of pre-stimulus activity on the subsequent perceptual awareness. As a confirmative analysis of the above pre-stimulus effects, we performed a whole-brain contrast of

seen versus unseen trials right at the stimulus onset. Again, the AIC, DLPFC, and ACC were found (Figure S4A). Although there were a few other regions (e.g., medial temporal gyrus, PreCu, and fusiform face areas) showing pre-stimulus effects (Figure S4A), they were unlikely to play a dominant role in access consciousness because they did not show significant post-stimulus effects of seen versus unseen trials. Lastly, while we performed neither temporal smoothing nor low-pass filtering during fMRI data preprocessing, we still sought to account for any unexpected effects that bring components of post-stimulus activity back to the stimulus onset. If this were true, it should apply to the

contrast of hit versus correction rejection in the supraliminal condition. Because no such pre-stimulus effect was found (Figure S4B), the above potential confound was considered unlikely.

The visual perception task also involved a DMN-DAT switch, but the DAT+ and DMN+ showed neither pre-stimulus nor post-stimulus effect (Figures 7F and 7G). Therefore, the DMN-DAT switch in this experimental paradigm may indicate an overall shift from internal to external awareness (triggered by the "view-response" task set) rather than a particular conscious content (face or not). As anticipated, post-stimulus effects were seen in the VAT+ and FPN+ (Figure S3), where the AIC, Thal, and ACC belonged to the VAT+, and the DLPFC was part of the FPN+.

DISCUSSION

The goal of this investigation was to identify the cortical gate for conscious access of sensory information. By manipulating both level of consciousness (first experiment) and content of consciousness (second experiment) in independent study protocols, we demonstrated that the AIC, situated at an intermediate position in the brain's functional hierarchy between primary sensory processing and higher-order processing, was the probable cortical site where conscious access to sensory information is gated. Imagery-related functional activation of the AIC was absent during the anesthetized state and was associated with an interruption of the ongoing DMN-DAT switches required for conscious processing. The causal role of the AIC in gating conscious access was confirmed in another set of experiments where conscious access could be predicted by the spontaneous pre-stimulus fluctuations of AIC activity near perceptual threshold.

AIC and failed conscious access in anesthesia

Conscious access is the apex of cognitive hierarchy and is supported by multiple stages of non-conscious processing (Dehaene and Changeux, 2011). Conscious access may fail in various conditions, such as when sensory stimuli are weak or corrupted (van Vugt et al., 2018), when attention is distracted, or when a person is sedated or has suffered a specific neurological injury (Alkire et al., 2008; Dehaene and Changeux, 2011; Mashour and Hudetz, 2018; Mashour et al., 2020a). Why does access consciousness fail in anesthesia? Our data showed that the dysfunction of the AIC renders higher-order systems inoperative, which may be the proximal cause of the disruption of conscious access by general anesthetics. This finding may address an important knowledge gap in why sensory inputs can be received but not perceived during anesthesia (i.e., why there is a preservation of low-level sensory processing concomitant with disruption of higher-order functions) (Boveroux et al., 2010; Davis et al., 2007; Gross et al., 2019; Mashour and Hudetz, 2018; Schroeder et al., 2016).

The AIC is a central component of the brain's salience network and ventral attention system (Fox et al., 2006; Menon, 2011; Seeley et al., 2007; Uddin, 2015), which is situated at an intermediate position between unimodal and transmodal areas along the brain's functional gradients (Huntenburg et al., 2018; Margulies et al., 2016; Wang et al., 2019). Anatomically, the AIC is composed of unique clusters of large spindle-shaped pyramidal

neurons in layer 5, called von Economo neurons (Allman et al., 2011). These neurons establish long-distance, fast relay of information throughout the cortex (Craig, 2009). Thus, the AIC has the neuroanatomical characteristics to support the global neuronal workspace posited to enable conscious access (Dehaene and Changeux, 2011; Mashour et al., 2020a; Michel, 2017). Functionally, the AIC has been recognized as a multifaceted region, playing a broad range of roles such as interoception, emotional awareness, visual and auditory awareness of the moment, attention, perceptual decision-making, cross-modal sensory processes, and cognitive control across many different domains (Craig, 2009; Sterzer and Kleinschmidt, 2010). The key role of the AIC seems to be identifying and prioritizing salient stimuli in the stream of continuous sensory information and sending signals to the systems responsible for the allocation of top-down attentional resources to the relevant sensory representations (Menon, 2011; Michel, 2017; Uddin, 2015). Furthermore, these activations of the AIC probably operate non-consciously or pre-consciously. For example, a subliminal no-go stimulus (i.e., a visual cue for inhibitory behavioral control) can activate the AIC and subsequently bias behavioral performance (e.g., a substantial reduction in response speed) (van Gaal et al., 2010). Once a no-go stimulus was consciously accessed, the AIC activity was amplified via top-down signaling from the PFC (Dehaene and Changeux, 2011; van Gaal et al., 2010). Therefore, the deactivation of the AIC during anesthesia suggests that the stimulus did not even reach preconscious level and therefore could not enable conscious processing.

The conclusion that the AIC is a critical cortical site for anesthetic-induced unresponsiveness concurs with the findings from a study that applied multisensory stimulation (i.e., auditory tones, words, and noxious pain stimuli) during propofol anesthesia (Mashour, 2016; Warnaby et al., 2016). The AIC was proposed as a "cortical gate" that, when deactivated during propofol-induced unresponsiveness (consistent with impaired conscious access), uncouples the DLPFC and the posterior parietal cortex despite the preservation of sensory-evoked responses (Warnaby et al., 2016).

Dysfunction of the AIC disables DMN-DAT switches

A recent study emphasized the role of anticorrelated DMN and DAT, which become isolated from the cycling patterns of brain networks (i.e., a "temporal circuit") during unconsciousness of diverse etiologies (Huang et al., 2020). Although there was evidence that the structured dynamics of network transitions were a defining feature of conscious processing, the neural region gating such transitions was not determined. In the present work, the AIC was identified across two experimental paradigms as a region whose deactivation was associated with a malfunction of DMN-DAT switches.

A series of prior studies demonstrated that the salience network, of which the AIC is a key node, plays a critical and causal role in the dynamic switching between the DMN and executive/attention (e.g., DAT) networks (Menon, 2011; Menon and Uddin, 2010; Uddin, 2015). A triple network model was proposed, in which the AIC is responsible for the brain network switch in order to facilitate access to attention and working memory resources when a salient event is detected (Menon,

2011). Additional evidence supporting the causal role of the AIC in mediating the DMN-DAT switch has been obtained, with multiple approaches including transcranial magnetic stimulation (Chen et al., 2013), Granger causality analysis (Chiong et al., 2013; Moran et al., 2013; Palaniyappan et al., 2013; Sridharan et al., 2008; Supekar and Menon, 2012; Uddin et al., 2011), dynamic causal modeling (Goulden et al., 2014), and Bayesian network learning (Li et al., 2018) of fMRI data in awake subjects. Based on our data in the context of anesthetic-induced unconsciousness, it seems plausible that the deactivation of the AIC may causally disrupt the DMN-DAT switches, resulting in a failure of conscious access. In other words, the AIC might serve as an arbitrator at the entrance of conscious access. Passing through this entrance, mental representations (e.g., mental imagery contents) are coupled with top-down attentional resources (activation of DAT) necessary for conscious access, while internally derived cognitive operations are suppressed (deactivation of DMN). With suppression of the AIC's activity during anesthesia, higher-order processing of sensory representations is presumably blocked or degraded, and conscious access becomes impossible.

Besides the AIC, we examined two other candidate brain regions (Ch4 and Thal) that may be involved in modulating large-scale brain network activity (Buckner and DiNicola, 2019; Halassa and Kastner, 2017). Although there were significant correlations for both the Ch4 and Thal with the DMN-DAT switch, multivariate linear regression analyses showed that the AIC was the only significant contributor to this network transition (Figure S5). We speculated that the Ch4 and Thal may be more closely related to arousal/vigilance regulation or level of consciousness (Brown et al., 2011; Liu et al., 2018; Turchi et al., 2018), instead of the cognitive control in mitigating neural resource competition as the AIC does. However, we recognize that the level of consciousness has a tight association with conscious contents (Bachmann and Hudetz, 2014; Northoff and Huang, 2017). For example, when the participants had normal wakefulness (e.g., during Base1), the mental imagery contents were consciously accessed, but this ability was modulated according to the level of vigilance and ultimately vanished during LOR. Nevertheless, the interplay between the level and content of consciousness is an important question that warrants further investigation.

AIC and perceptual awareness

We found that the spontaneous fluctuation of pre-stimulus AIC activity was predictive of conscious access for a fixed near-threshold visual stimulus. This finding is consistent with the pharmacological results and strengthens a casual role of the AIC in prioritizing salient stimuli to reach conscious access. The results support the hypothesis that visual consciousness is mediated by higher-order brain areas that are anterior to the visual cortex (Brascamp et al., 2015; Liu et al., 2019; van Vugt et al., 2018). Notably, van Vugt et al. (2018) studied the propagation of spiking activity elicited by weak visual stimuli in areas V1, V4, and DLPFC of monkeys. The investigators found that both perceived and unperceived stimuli caused activity in V1 and V4, but only the activity level in the DLPFC categorically predicted conscious access, implying that the DLPFC may lie at or beyond the stage that de-

termines the reporting threshold. The authors also reported evidence based on signal detection theory (SDT) for the separable influence of pre-stimulus activity on the subject's response bias (i.e., criterion) and sensitivity. For example, a lower criterion (e.g., higher hit and false alarm rates) was associated with increased pre-stimulus cortical excitation, whereas higher sensitivity (e.g., better discrimination of hits versus misses without influencing the false alarm rate) was related to an increase in the efficiency of bottom-up signal propagation (van Vugt et al., 2018). In line with this finding, we observed differential DLPFC activity for seen versus unseen trials, both pre-stimulus and post-stimulus (Figure 7E). Furthermore, we found that the cortical stage whose activity predicted conscious access of a near-threshold stimulus could be traced back to the upstream regions of DLPFC, including the AIC and its partners (ACC and Thal) in the salience network. This suggests that the pre-stimulus co-activation of the salience network may contribute to augmenting the post-stimulus signal allowing it to reach the DLPFC.

Admittedly, we were not able to ascertain whether the pre-stimulus effects are associated with criterion or sensitivity for the following reasons. In our experiment, the participants showed a highly conservative decision criterion in which the false alarm rate was very low (~4%). The limited number of near-threshold and false alarm trials was insufficient to form statistically distinct distributions to allow the dissociation of criterion and sensitivity in trials with low versus high pre-stimulus activity. A conservative criterion has often been observed in visual perception experiments (Iemi et al., 2017; Li et al., 2014; Limbach and Corballis, 2016; Podvalny et al., 2019; Wyart and Tallon-Baudry, 2008). The low false alarm rate in our data may also be because the perceptual threshold was determined as the stimulus presentation duration at which the hit rate of the face stimulus was close to chance, rather than the hit rate of scrambled face. In addition, the scrambled face in our experiment was not matched with the spatial frequency or contour contrast of the real face; the decision criterion would otherwise shift toward being less conservative, as seen in other experiments (Podvalny et al., 2019).

Besides the AIC and DLPFC, we observed pre-stimulus effect in the Thal and ACC. It is noteworthy that the AIC, ACC, and Thal are functionally connected within the salience network (Seeley et al., 2007; Uddin, 2015). There are widespread von Economo neurons within the salience network that participate in the efficient processing of information from different resources (Allman et al., 2011). However, the roles of the AIC and ACC are different. The AIC is a multimodal hub receiving and prioritizing multisensory inputs but has few motor outputs (Uddin, 2015). By contrast, the ACC serves functions more closely related to response selection, conflict monitoring, and attention for action through direct neuronal projections to the premotor and motor cortices.

With respect to the unique role of the DLPFC in access consciousness (Dehaene and Changeux, 2011; Mashour et al., 2020a; van Vugt et al., 2018) and combining our findings from both experiments, we postulate that the DLPFC may initiate a global broadcast of shared information available to mental-imagery-related processors (e.g., SMA, PPA, and PreCu in the pharmacological experiment) or to response selection hub (e.g., ACC in the psychological experiment). This ultimately led

to a conscious experience that was reportable either covertly (i.e., via volitional brain activity in the pharmacological experiment) or overtly (i.e., via motor response in the psychological experiment). On the other hand, the AIC (but not the DLPFC) was identified as the cortical failure site, showing deactivations during loss of consciousness in the pharmacological experiment. This coincided well with the findings identified during perceptual awareness in the psychological experiment (e.g., Figure 7E). Given the AIC's intermediate position along the brain's functional gradients (Huntenburg et al., 2018; Margulies et al., 2016; Wang et al., 2019) and its critical role in mediating brain network switches, we propose the AIC as the probable cortical site where conscious access to sensory information is gated, prior to the DLPFC's global broadcasting.

Limitations

Our study has a number of limitations. First, with current methodology, it is not possible to distinguish between phenomenal consciousness (referring to its experiential characteristics) and access consciousness (referring to broadcasting of sensory information to other cognitive processors), a distinction originally proposed by Block (2005). In our paradigm, auditory perception of the task command represents the primary sensory aspect, whereas broadcasting this to produce volitional mental imagery would be a sign of access—a form of covert post-sensory report. Our experimental design enables the study of conscious access rather than primary sensory perception. Both the mental imagery and visual masking tasks employed in our study recruit a wide range of post-perceptual cognitive resources, such as attention, motivation, working memory, and executive functions, which collectively contribute to access consciousness. Although the presence of access implies the presence of phenomenal experience, the converse may not be true. See Figure S6 for an illustration of the rationale for our experimental design and the interpretation with respect to the terminology and theoretical aspects of consciousness.

Second, different anesthetic agents may alter brain functions in different ways. Whether our results with respect to the failure of the AIC can be generalized to other classes of anesthetics including halogenated ethers, such as sevoflurane, or non-GABAergic drugs, such as ketamine, remains to be determined.

Third, the von Economo neurons in the AIC have been found only in species that are able to pass the standard mirror test for self-recognition such as elephants and macaques (Critchley and Seth, 2012; Evrard et al., 2012). Accordingly, the AIC has been hypothesized to be the seat of self-consciousness based on its role in binding information related to the self, interoception, and emotions to create a unitary "awareness of the self having the experience" (Craig, 2009). Our data do not allow us to elucidate how self-consciousness may contribute to initiating mental imagery or perceptual awareness in the AIC. Hence, more empirical work is needed to decipher the evolutionary and neural basis of self-consciousness and its uniquely human aspects (Mashour et al., 2020a).

Conclusions

The results suggest that dysfunction of the AIC during anesthesia disables DMN-DAT switching, resulting in the disruption of conscious access. Furthermore, pre-stimulus spontaneous fluctuations of AIC activity predict the subsequent conscious access of near-threshold stimuli. These findings support the hypothesis that the AIC is a central cortical region that gates conscious access.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Participants in pharmacological setting (Exp1)
  - Inclusion and exclusion criteria in Exp1
  - Participants in psychological setting (Exp2)
- METHOD DETAILS
  - Anesthetic agents in Exp1
  - Anesthetic administration and monitoring in Exp1
  - Experimental task during fMRI in Exp1
  - fMRI data acquisition in Exp1
  - Experimental stimuli in Exp2
  - Experimental task during fMRI in Exp2
  - fMRI data acquisition in Exp2
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - fMRI data preprocessing in Exp1
  - Whole brain general linear model analysis in Exp1
  - Cortical gradient analysis in Exp1
  - Definition of regions of interest in Exp1
  - Tracking time-locked modulation of large-scale co-activation patterns in Exp1
  - fMRI data analysis in Exp2
  - Statistical analysis in Exp1 and Exp2

AUTHOR CONTRIBUTIONS

Z.H. designed and conducted both the pharmacological and psychological experiments, analyzed the data, interpreted data, and wrote the article. V.T., P.E.V., E.L.J., and P.P. performed the anesthetic procedures. A.M.M. coordinated the experiment. G.A.M. interpreted data and wrote the article. A.G.H. conceived the project, designed the pharmacological experiment, interpreted data, and wrote the article.

REFERENCES

Alkire, M.T., Hudetz, A.G., and Tononi, G. (2008). Consciousness and anesthesia. Science *322*, 876–880.

Allman, J.M., Tetreault, N.A., Hakeem, A.Y., Manaye, K.F., Semendeferi, K., Erwin, J.M., Park, S., Goubert, V., and Hof, P.R. (2011). The von Economo neurons in the frontoinsular and anterior cingulate cortex. Ann. N Y Acad. Sci. *1225*, 59–71.

Anticevic, A., Cole, M.W., Murray, J.D., Corlett, P.R., Wang, X.J., and Krystal, J.H. (2012). The role of default network deactivation in cognition and disease. Trends Cogn. Sci. *16*, 584–592.

Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. Neurosci. Biobehav. Rev. *36*, 737–746.

Bachmann, T., and Hudetz, A.G. (2014). It is time to combine the two main traditions in the research on the neural correlates of consciousness: C=L×D. Front. Psychol. *5*, 1–13.

Bethlehem, R.A.I., Paquola, C., Seidlitz, J., Ronan, L., Bernhardt, B., Consortium, C.C.A.N., and Tsvetanov, K.A. (2020). Dispersion of functional gradients across the adult lifespan. Neuroimage *222*, 117299.

Block, N. (2005). Two neural correlates of consciousness. Trends Cogn. Sci. *9*, 46–52.

Boveroux, P., Vanhaudenhuyse, A., Bruno, M.-A.A., Noirhomme, Q., Lauwick, S., Luxen, A., Degueldre, C., Plenevaux, A., Schnakers, C., Phillips, C., et al. (2010). Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. Anesthesiology *113*, 1038–1053.

Brascamp, J., Blake, R., and Knapen, T. (2015). Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. Nat. Neurosci. *18*, 1672–1678.

Brown, E.N., Purdon, P.L., and Van Dort, C.J. (2011). General anesthesia and altered states of arousal: a systems neuroscience analysis. Annu. Rev. Neurosci. *34*, 601–628.

Brown, R., Lau, H., and LeDoux, J.E. (2019). Understanding the higher-order approach to consciousness. Trends Cogn. Sci. *23*, 754–768.

Buckner, R.L., and DiNicola, L.M. (2019). The brain's default network: updated anatomy, physiology and evolving insights. Nat. Rev. Neurosci. *20*, 593–608.

Chen, A.C., Oathes, D.J., Chang, C., Bradley, T., Zhou, Z.W., Williams, L.M., Glover, G.H., Deisseroth, K., and Etkin, A. (2013). Causal interactions between fronto-parietal central executive and default-mode networks in humans. Proc. Natl. Acad. Sci. USA *110*, 19944–19949.

Chiong, W., Wilson, S.M., D'Esposito, M., Kayser, A.S., Grossman, S.N., Poorzand, P., Seeley, W.W., Miller, B.L., and Rankin, K.P. (2013). The salience network causally influences default mode network activity during moral reasoning. Brain *136*, 1929–1941.

Cox, R.W., Chen, G., Glen, D.R., Reynolds, R.C., and Taylor, P.A. (2017). FMRI clustering in AFNI: False-positive rates redux. Brain Connect. *7*, 152–171.

Craig, A.D. (2009). How do you feel—now? The anterior insula and human awareness. Nat. Rev. Neurosci. *10*, 59–70.

Critchley, H., and Seth, A. (2012). Will studies of macaque insula reveal the neural mechanisms of self-awareness? Neuron *74*, 423–426.

Davis, M.H., Coleman, M.R., Absalom, A.R., Rodd, J.M., Johnsrude, I.S., Matta, B.F., Owen, A.M., and Menon, D.K. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. Proc. Natl. Acad. Sci. USA *104*, 16032–16037.

Dehaene, S., and Changeux, J.-P.P. (2011). Experimental and theoretical approaches to conscious processing. Neuron *70*, 200–227.

Evrard, H.C., Forro, T., and Logothetis, N.K. (2012). Von Economo neurons in the anterior insula of the macaque monkey. Neuron *74*, 482–489.

Fiset, P., Paus, T., Daloze, T., Plourde, G., Meuret, P., Bonhomme, V., Hajj-Ali, N., Backman, S.B., and Evans, A.C. (1999). Brain mechanisms of propofol-induced loss of consciousness in humans: a positron emission tomographic study. J. Neurosci. *19*, 5506–5513.

Fox, M.D., Corbetta, M., Snyder, A.Z., Vincent, J.L., and Raichle, M.E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. Proc. Natl. Acad. Sci. USA *103*, 10046–10051.

Goulden, N., Khusnulina, A., Davis, N.J., Bracewell, R.M., Bokde, A.L., McNulty, J.P., and Mullins, P.G. (2014). The salience network is responsible for switching between the default mode network and the central executive network: replication from DCM. Neuroimage *99*, 180–190.

Gross, W.L., Lauer, K.K., Liu, X., Roberts, C.J., Liu, S., Gollapudy, S., Binder, J.R., Li, S.J., and Hudetz, A.G. (2019). Propofol sedation alters perceptual and cognitive functions in healthy volunteers as revealed by functional magnetic resonance imaging. Anesthesiology *131*, 254–265.

Hairston, W.D., and Maldjian, J.A. (2009). An adaptive staircase procedure for the E-Prime programming environment. Comput. Methods Programs Biomed. *93*, 104–108.

Halassa, M.M., and Kastner, S. (2017). Thalamic functions in distributed cognitive control. Nat. Neurosci. *20*, 1669–1679.

Hong, S.J., Vos de Wael, R., Bethlehem, R.A.I., Lariviere, S., Paquola, C., Valk, S.L., Milham, M.P., Di Martino, A., Margulies, D.S., Smallwood, J., and Bernhardt, B.C. (2019). Atypical functional connectome hierarchy in autism. Nat. Commun. *10*, 1022.

Huang, Z., Zhang, J., Longtin, A., Dumont, G., Duncan, N.W., Pokorny, J., Qin, P., Dai, R., Ferri, F., Weng, X., and Northoff, G. (2017). Is there a nonadditive interaction between spontaneous and evoked activity? Phase-dependence and its relation to the temporal structure of scale-free brain activity. Cereb. Cortex *27*, 1037–1059.

Huang, Z., Vlisides, P.E., Tarnal, V.C., Janke, E.L., Keefe, K.M., Collins, M.M., McKinney, A.M., Picton, P., Harris, R.E., Mashour, G.A., and Hudetz, A.G. (2018a). Brain imaging reveals covert consciousness during behavioral unresponsiveness induced by propofol. Sci. Rep. *8*, 13195.

Huang, Z., Liu, X., Mashour, G.A., and Hudetz, A.G. (2018b). Timescales of intrinsic BOLD signal dynamics and functional connectivity in pharmacologic and neuropathologic states of unconsciousness. J. Neurosci. *38*, 2304–2317.

Huang, Z., Zhang, J., Wu, J., Mashour, G.A., and Hudetz, A.G. (2020). Temporal circuit of macroscale dynamic brain activity supports human consciousness. Sci. Adv. *6*, eaaz0087.

Hudetz, A.G. (2006). Suppressing consciousness: Mechanisms of general anesthesia. Semin. Anesth. Perioper. Med. Pain *25*, 196–204.

Huntenburg, J.M., Bazin, P.L., and Margulies, D.S. (2018). Large-scale gradients in human cortical organization. Trends Cogn. Sci. *22*, 21–31.

Iemi, L., Chaumon, M., Crouzet, S.M., and Busch, N.A. (2017). Spontaneous neural oscillations bias perception by modulating baseline excitability. J. Neurosci. *37*, 807–819.

LaConte, S.M., Peltier, S.J., and Hu, X.P. (2007). Real-time fMRI using brain-state classification. Hum. Brain Mapp. *28*, 1033–1044.

Li, Q., Hill, Z., and He, B.J. (2014). Spatiotemporal dissociation of brain activity underlying subjective awareness, objective performance and confidence. J. Neurosci. *34*, 4382–4395.

Li, R., Zhang, S., Yin, S., Ren, W., He, R., and Li, J. (2018). The fronto-insular cortex causally mediates the default-mode and central-executive networks to contribute to individual cognitive performance in healthy elderly. Hum. Brain Mapp. *39*, 4302–4311.

Limbach, K., and Corballis, P.M. (2016). Prestimulus alpha power influences response criterion in a detection task. Psychophysiology *53*, 1154–1164.

Liu, X., Lauer, K.K., Ward, B.D., Li, S.J., and Hudetz, A.G. (2013). Differential effects of deep sedation with propofol on the specific and nonspecific thalamocortical systems: a functional magnetic resonance imaging study. Anesthesiology *118*, 59–69.

Liu, X., de Zwart, J.A., Schölvinck, M.L., Chang, C., Ye, F.Q., Leopold, D.A., and Duyn, J.H. (2018). Subcortical evidence for a contribution of arousal to fMRI studies of brain activity. Nat. Commun. *9*, 395.

Liu, S., Yu, Q., Tse, P.U., and Cavanagh, P. (2019). Neural correlates of the conscious perception of visual location lie outside visual cortex. Curr. Biol. *29*, 4036–4044.e4.

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., et al. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. USA *113*, 12574–12579.

Marsh, B., White, M., Morton, N., and Kenny, G.N.C. (1991). Pharmacokinetic model driven infusion of propofol in children. Br. J. Anaesth. *67*, 41–48.

Mashour, G.A. (2016). Anesthetizing the self: The neurobiology of humbug. Anesthesiology *124*, 747–749.

Mashour, G.A., and Alkire, M.T. (2013). Consciousness, anesthesia, and the thalamocortical system. Anesthesiology *118*, 13–15.

Mashour, G.A., and Hudetz, A.G. (2018). Neural correlates of unconsciousness in large-scale brain networks. Trends Neurosci. *41*, 150–160.

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

Mashour, G.A., Palanca, B.J., Basner, M., Li, D., Wang, W., Blain-Moraes, S., Lin, N., Maier, K., Muench, M., Tarnal, V., et al. (2020b). Recovery of consciousness and cognition after general anesthesia in humans. BioRxiv, 2020.05.28.121269.

Mckeown, B., Strawson, W.H., Wang, H.T., Karapanagiotidis, T., Vos de Wael, R., Benkarim, O., Turnbull, A., Margulies, D., Jefferies, E., McCall, C., et al. (2020). The relationship between individual variation in macroscale functional gradients and distinct aspects of ongoing thought. Neuroimage *220*, 117072.

Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. Trends Cogn. Sci. *15*, 483–506.

Menon, V., and Uddin, L.Q. (2010). Saliency, switching, attention and control: a network model of insula function. Brain Struct. Funct. *214*, 655–667.

Michel, M. (2017). A role for the anterior insular cortex in the global neuronal workspace model of consciousness. Conscious. Cogn. *49*, 333–346.

Monti, M.M., Vanhaudenhuyse, A., Coleman, M.R., Boly, M., Pickard, J.D., Tshibanda, L., Owen, A.M., and Laureys, S. (2010). Willful modulation of brain activity in disorders of consciousness. N. Engl. J. Med. *362*, 579–589.

Moran, L.V., Tagamets, M.A., Sampath, H., O'Donnell, A., Stein, E.A., Kochunov, P., and Hong, L.E. (2013). Disruption of anterior insula modulation of large-scale brain networks in schizophrenia. Biol. Psychiatry *74*, 467–474.

Murphy, K., and Fox, M.D. (2017). Towards a consensus regarding global signal regression for resting state functional connectivity MRI. Neuroimage *154*, 169–173.

Northoff, G., and Huang, Z. (2017). How do the brain's time and space mediate consciousness and its different dimensions? Temporo-spatial theory of consciousness (TTC). Neurosci. Biobehav. Rev. *80*, 630–645.

Owen, A.M., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S., and Pickard, J.D. (2006). Detecting awareness in the vegetative state. Science *313*, 1402.

Palaniyappan, L., Simmonite, M., White, T.P., Liddle, E.B., and Liddle, P.F. (2013). Neural primacy of the salience processing system in schizophrenia. Neuron *79*, 814–828.

Paquola, C., Bethlehem, R.A., Seidlitz, J., Wagstyl, K., Romero-Garcia, R., Whitaker, K.J., Vos de Wael, R., Williams, G.B., Vértes, P.E., Margulies, D.S., et al.; NSPN Consortium (2019). Shifts in myeloarchitecture characterise adolescent development of cortical gradients. eLife *8*, e50482.

Podvalny, E., Flounders, M.W., King, L.E., Holroyd, T., and He, B.J. (2019). A dual role of prestimulus spontaneous neural activity in visual object recognition. Nat. Commun. *10*, 3910.

Raichle, M.E. (2015). The brain's default mode network. Annu. Rev. Neurosci. *38*, 433–447.

Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., and Yeo, B.T.T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cereb. Cortex *28*, 3095–3114.

Schroeder, K.E., Irwin, Z.T., Gaidica, M., Nicole Bentley, J., Patil, P.G., Mashour, G.A., and Chestek, C.A. (2016). Disruption of corticocortical information transfer during ketamine anesthesia in the primate brain. Neuroimage *134*, 459–465.

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., and Greicius, M.D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. *27*, 2349–2356.

Shafer, S. (1996). STANPUMP User's Manual (STANPUMP). http://opentci.org/code/stanpump.

Sherman, S.M. (2016). Thalamus plays a central role in ongoing cortical functioning. Nat. Neurosci. *19*, 533–541.

Sridharan, D., Levitin, D.J., and Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. Proc. Natl. Acad. Sci. USA *105*, 12569–12574.

Sterzer, P., and Kleinschmidt, A. (2010). Anterior insula activations in perceptual paradigms: often observed but barely understood. Brain Struct. Funct. *214*, 611–622.

Supekar, K., and Menon, V. (2012). Developmental maturation of dynamic causal control signals in higher-order cognition: a neurocognitive network model. PLoS Comput. Biol. *8*, e1002374.

Suzuki, M., and Larkum, M.E. (2020). General Anesthesia Decouples Cortical Pyramidal Neurons. Cell *180*, 666–676.e13.

Tanabe, S., Huang, Z., Zhang, J., Chen, Y., Fogel, S., Doyon, J., Wu, J., Xu, J., Zhang, J., Qin, P., et al. (2020). Altered global brain signal during physiologic, pharmacologic, and pathologic states of unconsciousness in humans and rats. Anesthesiology *132*, 1392–1406.

Turchi, J., Chang, C., Ye, F.Q., Russ, B.E., Yu, D.K., Cortes, C.R., Monosov, I.E., Duyn, J.H., and Leopold, D.A. (2018). The basal forebrain regulates global resting-state fMRI fluctuations. Neuron *97*, 940–952.e4.

Uddin, L.Q. (2015). Salience processing and insular cortical function and dysfunction. Nat. Rev. Neurosci. *16*, 55–61.

Uddin, L.Q., Supekar, K.S., Ryali, S., and Menon, V. (2011). Dynamic reconfiguration of structural and functional connectivity across core neurocognitive brain networks with development. J. Neurosci. *31*, 18578–18589.

van Gaal, S., Ridderinkhof, K.R., Scholte, H.S., and Lamme, V.A.F. (2010). Unconscious activation of the prefrontal no-go network. J. Neurosci. *30*, 4143–4150.

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P.R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. Science *360*, 537–542.

Vos de Wael, R., Benkarim, O., Paquola, C., Lariviere, S., Royer, J., Tavakol, S., Xu, T., Hong, S.J., Langs, G., Valk, S., et al. (2020). BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. Commun. Biol. *3*, 103.

Wang, P., Kong, R., Kong, X., Liégeois, R., Orban, C., Deco, G., van den Heuvel, M.P., and Thomas Yeo, B.T. (2019). Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain. Sci. Adv. *5*, eaat7854.

Warnaby, C.E., Seretny, M., Ní Mhuircheartaigh, R., Rogers, R., Jbabdi, S., Sleigh, J., and Tracey, I. (2016). Anesthesia-induced suppression of human dorsal anterior insula responsivity at loss of volitional behavioral response. Anesthesiology *124*, 766–778.

Wyart, V., and Tallon-Baudry, C. (2008). Neural dissociation between visual awareness and spatial attention. J. Neurosci. *28*, 2667–2679.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. *106*, 1125–1165.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Propofol (2,6 diisopropylphenol)	Fresenius Kabi USA, LLC	N/A
Software and algorithms		
MATLAB R2017b	https://www.mathworks.com	RRID:SCR_001622
STANPUMP (Shafer, 1996)	http://opentci.org/code/stanpump	N/A
E-Prime 3.0	https://www.pstnet.com/eprime.cfm	RRID:SCR_009567
AcqKnowledge Software (V5.0)	https://www.biopac.com/product/acqknowledge-software/	RRID:SCR_014279
Analysis of Functional NeuroImages (AFNI)	https://afni.nimh.nih.gov/afni/	RRID:SCR_005927
BrainSpace Toolbox	https://brainspace.readthedocs.io/en/latest/	N/A
Adobe Photoshop	https://www.adobe.com/products/photoshop.html	RRID:SCR_014199

RESOURCE AVAILABILITY

Lead contact
Requests for further information and resources should be directed to and will be fulfilled by the Lead Contact, Anthony G. Hudetz (ahudetz@med.umich.edu).

Materials availability
This study did not generate reagents.

Data and code availability
Publicly available software used for analyses is listed in the Key resources table. Code used in this study will be shared upon request. Access to data by qualified investigators (i.e., affiliated with accredited academic and research institutions) are subject to scientific and ethical review and must comply with the National Institutes of Health (NIH) regulations. Completion of a material transfer agreement signed by an institutional official will be required in order to access the data.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants in pharmacological setting (Exp1)
The University of Michigan Institutional Review Board (IRB) approved the experimental protocol. All methods were performed in accordance with the relevant guidelines and regulations. Following careful discussion and written informed consent, twenty-six participants (ages between 19-34 years old; 13 women) were recruited. Strict confidentiality was maintained throughout. Participants were assigned a code number following their first contact in the protocol. This number was used throughout the experiment and was the only identifier on specimen samples, behavioral and physiological archival data, and magnetic resonance (MR) scans.

Inclusion and exclusion criteria in Exp1
Right-handed healthy participants of American Society of Anesthesiologists physical status 1, aged 18-40, with a body mass index < 30, who were experienced with racquet sports (at least 30 times over their lifetime) were eligible for inclusion. Subjects were excluded from participation if they did not speak English; had any contraindication to MRI scanning; possible pregnancy, extreme obesity, metallic substances in the body, claustrophobia, anxiety, or cardiopulmonary disease; had a history of neurological, cardiovascular, or pulmonary illness; significant head injury with loss of consciousness; learning disability or other developmental disorder; sleep apnea or any severe snoring history; or sensory/motor loss sufficient to interfere with performance of the study, gastresophageal reflux disease, unwilling to abstain from alcohol use for 24 hours prior to their scheduled MRI study visit, history of drug use or a positive drug screen, tattoos on the head or neck region; had a history of allergic reaction to eggs; had an intracranial structural abnormality on T1-weighted MRI scans or experienced physical discomfort during fMRI scanning.

Participants in psychological setting (Exp2)

Twenty-five right-handed volunteers participated in this study (ages between 18-29 years old; 12 women). Six of them were excluded due to falling asleep, excessive head movement, or extremely high or low hit rates (> 85% or < 15%; 2 SDs) of near-threshold face presentation during fMRI scanning. All participants were all university students with normal or corrected-to-normal vision, and had no history of psychiatric or neurological disorders. Written informed consent was obtained from each subject before the experiment. The study was approved by the ethics committee of the Center for Cognition and Brain Disorders (CCBD), Hangzhou Normal University, Hangzhou, Zhejiang, China.

METHOD DETAILS

Anesthetic agents in Exp1

Propofol was our reference drug because it has been the most widely-used agent in human fMRI studies of anesthetic effects. The advantage of propofol is that it exerts minimal effects on cerebral hemodynamics (Fiset et al., 1999); because it preserves flow-meta-bolism coupling in the cerebral vasculature, it minimizes confound of the fMRI interpretation. Propofol suppresses neuronal activity mainly through an enhancement of GABA-A receptor-mediated inhibition thus modulating widespread targets throughout the brain (Alkire et al., 2008). In terms of safety in healthy volunteers, a multicenter, 30-participant study demonstrated no adverse effects of surgical anesthesia (in the absence of surgery), with cognition returning to baseline 3-hours after emergence from a prolonged anesthetic, without signs of disrupted arousal states in the following days, suggesting the healthy human brain is resilient even to deep anesthesia (Mashour et al., 2020b).

Anesthetic administration and monitoring in Exp1

All subjects fasted for 8 hours before the study. On the day of the experiment, an attending anesthesiologist completed a pre-operative assessment and physical examination. Two fully trained anesthesiologists were physically present for the entire duration of the experiment. An intravenous cannula was placed after a subcutaneous injection of lidocaine (0.5 mL of 1%) used as local anesthetic. Spontaneous respiration, end-tidal CO_2, heart rate, pulse oximetry, and electrocardiogram were continuously monitored during the experiment. Noninvasive arterial pressure was measured with MR-compatible automatic monitor. Supplemental oxygen (2L/min via nasal cannula) was used for all subjects. The propofol administration was achieved by target-controlled IV bolus and constant rate infusion. The bolus dose, infusion rate and infusion duration for each target effect-site concentration (ESC) and for each participant were pre-determined based on a pharmacokinetic model (Marsh et al., 1991) developed for target-controlled propofol infusion and implemented in software (STANPUMP; Shafer, 1996). The dosing (bolus + infusion) was incremented at every 5 minutes until the final target was reached. The incremental dosing (0.4 μg/ml) was used to titrate the anesthetic level to the point of loss of behavioral responsiveness (LOR). The initial target ESC was 0.4 μg/ml in 14 participants and 1.0 μg/ml in 12 participants. The final target concentration was 2.4 μg/ml (in 6 participants from our previous study (Huang et al., 2018a)) or one increment above that first resulted in LOR (in 20 participants). For example, if a participant showed LOR at 2.0 μg/ml target concentration, we used 2.4 μg/ml as final target. The reason was that a relatively low dose of propofol (ESC close to LOR) often induces agitation resulting in frequent or large head movements, and a slightly higher dose (e.g., one increment above) could reduce head motion effects in fMRI data. The final target was maintained for 21.6 minutes on average (±SD = 10.2 minutes). After that, the infusion was terminated to allow spontaneous emergence. The ESC of LOR in our current studied sample (n = 26) was 1.8 ± 0.6 μg/ml (mean ± SD).

Experimental task during fMRI in Exp1

Mental imagery and motor response tasks were studied before, during and after stepwise propofol infusion. Participants were asked to perform three imagery tasks (tennis, navigation and hand squeeze) plus a motor response task (actual hand squeeze). For tennis imagery, they were instructed to imagine standing still on a tennis court and to swing an arm to hit the ball back and forth to an imagined instructor. For navigation imagery, participants were instructed to imagine navigating the streets of a familiar city or to imagine walking from room to room in their home and to visualize all that they would see if they were there (Owen et al., 2006). In the squeeze imagery task, participants were instructed to imagine squeezing an MRI compatible grip dynamometer (a rubber ball). In the motor response task, participants were instructed to actually grip the rubber ball (task following the squeeze imagery). A pseudo-randomized (Latin square) block design was applied, in which 15 s periods of tennis (and navigation) imagery, and 10 s periods of squeeze imagery with a hand squeeze within 5 s after hearing the instruction, alternated with 15 s of rest. The entire scan included 180 rest–imagery cycles (60 cycles per condition). The beginning of each trial was cued with the spoken word "tennis imagery," "navigation imagery," "squeeze imagery," or "action," and the rest period was cued with the word "relax."

The verbal instructions were programmed using E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA) and delivered via an audiovisual stimulus presentation system designed for an MRI environment. The volume of the headphones was adjusted for subject comfort. Behavioral responses were measured in mmHg of air pressure during squeezing the rubber ball, using BIOPAC (https://www.biopac.com) MP160 system with AcqKnowledge software (V5.0). By comparing the timing of "action" instructions and the actual motor response during and after propofol infusion, the periods during which a subject retained responsiveness (PreLOR), loss of responsiveness (LOR), and recovery of responsiveness (ROR) were determined. The offset of PreLOR, onset of LOR, offset of LOR, and onset of ROR were defined as the times of the last successful response of squeezing, the first failure to squeeze, the last

failure to squeeze, and the first successful response of squeezing after LOR, respectively. The data during the transition periods between PreLOR and LOR and between LOR and ROR were not included in the analysis because the temporal resolution of the behavioral assessment (instruction to squeeze the ball) was 90 s on average, within which the behavioral responsiveness was uncertain.

fMRI data acquisition in Exp1

Data were acquired at University of Michigan Hospital using a 3T Philips scanner with a standard 32-channel transmit/receive head coil. Before fMRI scans, T1 weighted spoiled gradient recalled echo (SPGR) images was acquired for high spatial resolution of anatomical images with parameters: 170 sagittal slices, 1.0mm thickness (no gap), TR = 8.1 s, TE = 3.7ms, flip angle = 8°, FOV = 24cm, image matrix 256 × 256. Functional images over the whole brain were acquired by a gradient-echo EPI pulse sequence with parameters: 28 slices, TR/TE = 800/25ms by multiband acquisition, MB factor = 4, slice thickness = 4mm, in-plane resolution = 3.4 × 3.4mm; field of view (FOV) = 220mm, flip angle = 76°, image matrix: 64 × 64. Six participants were scanned with slightly different parameters before MRI hardware upgradation (21 slices, TR/TE = 800/25ms, MB factor = 3, slice thickness = 6mm). All participants were asked to lay at rest with eyes closed in the scanner for the first 10-min (Rest1) and the last 10-min (Rest2) resting-state scan. They were asked not to move and to stay awake. Verbal instructions were presented through earphones. Four task fMRI runs were conducted including 15-min wakeful baseline (Base1), during (30-min) and after (30-min) propofol infusion, and another 15-min recovery baseline (Base2). There were five participants who regained behavioral responsiveness after the second 30-min fMRI run, four participants' ROR data were too short (< 4min) to yield reliable results, and one participant did not complete the study due to excessive body movement during ROR and whose data for Base2 and Rest2 were not acquired. In sum, there were 26 valid data for Rest1, Base1, PreLOR and LOR, 25 valid data for Base2 and Rest2, and 16 valid data for ROR.

Experimental stimuli in Exp2

Participants were flashed with either a face or a scrambled-face, which were then masked by a high-contrast image. They were instructed to report whether they had seen a face or not. A face image was edited using Adobe Photoshop (San Jose, CA) to crop it into an elliptical shape without external features (hair, ears) and create a uniform gray background. The face image was blurred, and brightness and contrast were reduced such that every subject could reach a threshold with duration longer than 16.7ms – the limit of the projector refresh rate (60Hz). A scrambled-face was created by randomly rearranging the face image, keeping an identical rounded frame to prevent subjects employing motion detection strategies to discriminate face from non-face. The high-contrast image (mask) was presented for 400ms immediately after the target being displayed.

We determined the supraliminal (above threshold) and near-threshold stimulus presentations by manipulating the duration of the target stimuli. For the supraliminal condition, a 200ms target duration was used. Individual thresholds for discriminating a face from a scrambled-face were determined by an adaptive staircase procedure (Hairston and Maldjian, 2009). This was performed inside the scanner before fMRI scanning. The threshold was defined as the point (i.e., stimulus presentation duration) at which the hit rate of a face was at chance. Three staircases ran independently, starting at target stimulus durations of 16.7, 50, and 66.7ms respectively (step size: 16.7ms). The duration increased one step after two consecutive incorrect responses, and decreased one step after each correct response. Each staircase terminated after 16 reversals, and an average was calculated from the last 5 reversal values. An overall average was then determined for the three staircases and rounded up to the nearest value compatible with the vertical scan rate of the projector (i.e., 16.7ms increments). The threshold across subjects (n = 19) was as follows: seventeen subjects, 33.3ms; two subjects, 66.7ms.

Dark adaptation (at least 20min) was conducted for all subjects before any behavioral data were collected. Subjects were unaware of the backward masking nature of the study. They were only informed that they would briefly see a photograph of a face and a noise picture, and then be required to make a button response. The individual threshold for the face was used later in the fMRI scanning for both the near-threshold face and scrambled-face presentations.

Experimental task during fMRI in Exp2

We adopted a sparse event-related fMRI with a two-by-two factorial design (face versus scrambled-face) × (supraliminal versus near-threshold). Each participant had to complete a total of 160 trials distributed across 8 functional runs (20 trials/run; ~11min/run). Each stimulus was repeated for 40 trials. All stimuli were counterbalanced in each run so that each run contained 5 trials per stimulus, assigned pseudo-randomly.

Each trial started with a brief flash of face or scrambled-face (either supraliminal or near-threshold) followed by a mask. Participants were instructed to view the stimuli but not respond until a red fixation cross prompt appeared on the screen. The variable delay period (white fixation cross) between stimulus and response prompt was from 5.4 s to 8.4 s with 1.5 s step size, plus a compensative time (e.g., 167ms) for near-threshold presentation (e.g., 33ms) in order to match with the total duration of supraliminal condition (200ms). Participants were instructed to respond using their index finger to indicate "seen" and their middle finger to indicate "unseen" on their right hand via an MRI compatible response box. The red fixation-cross returned to white again after the button response. A response time longer than 4.5 s was considered as a miss trial. After the button response, a long rest period with white fixation-cross was presented, varying unpredictably in length between 16.5 and 30 s for separate trials (19.5 s mean duration; 1.5 s steps; following an exponential distribution). Hence, the inter-trial-intervals (ITIs) between two stimuli were from 27 to 43.5 s. The benefit of such long ITIs is that they provided sufficient time to include the evoked positive BOLD response as well as its undershoot and then return to the

ongoing "baseline" level (Huang et al., 2017). They also avoid potential nonlinearities associated with overlapping hemodynamic responses between preceding and subsequent trials. All stimuli were programmed using E-Prime (Psychology Software Tools, Pittsburgh, PA) and delivered via an audiovisual stimulus presentation system designed for an MRI environment.

fMRI data acquisition in Exp2

Data were acquired at Hangzhou Normal University using a 3T GE scanner with a standard 8-channel transmit/receive head coil. Functional images over the whole brain were acquired by a gradient-echo EPI pulse sequence with parameters: 32 slices, TR/TE = 1500/30ms, slice thickness = 4mm, in-plane resolution = 3.4 × 3.4mm; field of view (FOV) = 220mm, flip angle = 72°, image matrix: 64 × 64. High-resolution anatomical images (180 sagittal slices, 1.0mm thickness, TR = 8.2 s, flip angle = 12°, FOV = 24cm, image matrix 256 × 256) were acquired at the end of the experiment. Before each fMRI run, participants were instructed to look at the crosshair, remain still, stay awake and not move their heads. The button response was monitored during the whole experiment to ensure participants' cooperation and alertness. For instance, a scan was stopped if no button response was detected in three successive trials. The participant was then asked if she/he had fallen asleep or had any other issue during scanning. After that, if resolved, the scan was repeated.

QUANTIFICATION AND STATISTICAL ANALYSIS

fMRI data preprocessing in Exp1

Preprocessing steps were implemented in AFNI (https://afni.nimh.nih.gov/afni). 1) Slice timing correction; 2) Rigid head motion correction/realignment within and across runs; frame-wise displacement (FD) of head motion was calculated using frame-wise Euclidean Norm (square root of the sum squares) of the six-dimension motion derivatives. A frame and its each previous frame were tagged as zeros (ones, otherwise) if the given frame's derivative value has a Euclidean Norm above FD = 0.5 mm; 3) Coregistration with high-resolution anatomical images; 4) Spatial normalization into Talaraich stereotactic space and resampling to 3 × 3 × 3 mm^3; 5) Using AFNI's function 3dTproject, the time-censored data were high-pass filtered above 0.008 Hz. At the same time, various undesired components (e.g., physiological estimates, motion parameters) were removed via linear regression. The undesired components included linear and nonlinear drift, time series of head motion and its temporal derivative, and mean time series from the white matter and cerebrospinal fluid; 6) Spatial smoothing with 6 mm full-width at half-maximum isotropic Gaussian kernel; 7) The time-course per voxel of each run was normalized to zero mean and unit variance, accounting for differences in variance of non-neural origin (e.g., distance from head coil). Unless otherwise stated, global signal regression (GSR) procedure was not applied.

Whole brain general linear model analysis in Exp1

Four types of events corresponded to the three mental imagery tasks (tennis, navigation and squeeze), and motor response (action) were defined for each of the five conditions (Base1, PreLOR, LOR, ROR and Base2). Events for each regressor was modeled and estimated by convolving onset times with a canonical hemodynamic response function using a BLOCK-model of the 3dDeconvolve function in AFNI. The event durations in the BLOCK-models were 15 s for tennis, 15 s for navigation, 10 s for squeeze, and 2 s for action. Estimated activation amplitude (beta value) were calculated at the voxel level across the whole brain after censoring out the frames with FD > 0.5 mm. The above produce was applied in a separate general linear model to estimate verbal instruction induced activation, in which only one type of event with 2 s duration was modeled by collapsing all onset times of verbal instructions (i.e., "tennis imagery," "navigation imagery," "squeeze imagery," "action" and "relax"). In total, five activation maps (tennis, navigation, squeeze, action and instruction) were generated for each condition per subject.

Cortical gradient analysis in Exp1

Following preprocessing, the fMRI time courses were extracted from 400 cortical areas according to a well-established brain parcellation scheme (Schaefer et al., 2018). A connectivity matrix was then calculated using Pearson correlation resulting in a 400x400 connectivity matrix for each participant and each condition (e.g., LOR). These individual connectivity matrices were then averaged to calculate a group-average per condition. Cortical gradients were computed using the BrainSpace toolbox as implemented in MATLAB (Vos de Wael et al., 2020). As in previous work (Hong et al., 2019; Margulies et al., 2016; Mckeown et al., 2020), we z-transformed and thresholded the connectivity matrix, leaving only the top 10% of weighted connections per row, and calculated a normalized angle affinity matrix that captures the similarity of connectivity profiles between cortical areas. We then applied a diffusion map embedding algorithm to identify principal gradient components, which estimates a low-dimensional embedding from a high-dimensional connectivity matrix (Margulies et al., 2016). The algorithm is controlled by two parameters α and t, where α controls the influence of the density of sampling points on the manifold ($\alpha = 0$, maximal influence; $\alpha = 1$, no influence) and t controls the scale of eigenvalues of the diffusion operator. We followed recommendations and fixed α at 0.5 and t at 0, a choice that retains the global relations between data points in the embedded space (Bethlehem et al., 2020; Hong et al., 2019; Margulies et al., 2016; Mckeown et al., 2020; Paquola et al., 2019; Vos de Wael et al., 2020). The group-level gradient solutions were aligned using Procrustes rotation to a subsample of the HCP dataset (n = 217) that is openly available in the Brainspace toolbox (Vos de Wael et al., 2020). This alignment step improves the stability of gradient estimation and maximizes the comparability of solutions to those from existing literature. Using identical parameters, individual-level gradients were then calculated for each condition. In addition, the

gradient eigenvector loading values were extracted from seven pre-defined functional networks (Yeo et al., 2011) in order to depict the cortical gradient organization at the network level.

Definition of regions of interest in Exp1

Regions of interest (ROIs) were defined by the group-level z-maps of Base1. Specifically, spatially segregated clusters (nearest-neighbor of faces touching; cluster size > 80 voxels) was obtained by adjusting the threshold (z-values from −4 to 6) on the averaged z-map across tennis, navigation, squeeze and action. These spatial clusters were identified as the supplementary motor area (SMA), precuneus (PreCu), parahippocampal place area (PPA), medial prefrontal cortex (MPFC), posterior cingulate cortex (PCC), visual cortex (Vis), postcentral gyrus (PostC), anterior insular cortex (AIC) and dorsal lateral prefrontal cortex (DLPFC). The auditory cortex (A1) and primary motor cortex (M1) were defined by the peak z-value clusters of the group-level z-maps of instruction and action, respectively. The bilateral thalamus (Thal) was defined using AFNI anatomical parcellations from the TT_caez_ml_18+tlrc atlas (regions 77-78), and the basal forebrain Ch4 was from the TT_caez_mpm_22+tlrc atlas (region 142). To account for individual variability, the above ROIs was clipped by computing the conjunctions between group-level ROIs and the corresponding first-level activation maps (p < 0.001, uncorrected) on a subject-by-subject basis.

Tracking time-locked modulation of large-scale co-activation patterns in Exp1

We calculated the spatial similarity between the signal intensity of each fMRI volume and eight pre-defined centroids of co-activation patterns (CAPs) derived from our previous study (Huang et al., 2020). The CAPs were classified as default-mode network (DMN+), dorsal attention network (DAT+), frontoparietal network (FPN+), sensory and motor network (SMN+), visual network (VIS+), ventral attention network (VAT+), and global network of activation and deactivation (GN+ and GN-). The eight CAPs were divided into four pairs of "mirror" motifs, with a strong negative spatial similarity. For instance, the DMN+ was accompanied by co-deactivation of DAT (DAT-), and vice versa for DAT+ (DMN-). Accordingly, for each condition and for each participant, the preprocessed fMRI data were transformed into eight CAP spatial similarity time courses. Next, the time courses time-locked to the onset of mental imagery tasks were averaged. The area under the curve (AUC of 4 s – 20 s following verbal instruction corrected by pre-stimulus baseline period −4.0 s – −0.8 s) was defined as the CAP modulation index to quantify the temporal gain of CAP transitioning.

fMRI data analysis in Exp2

The preprocessing steps were the same with those aforementioned. After that, trials were sorted into fours bins: hit, miss, false alarm, and correct rejection according to the participant's response for supraliminal and near-threshold conditions, respectively. As the numbers of trials in supraliminal miss, supraliminal false alarm, and near-threshold false alarm were very low, those trials were discarded. Hence, five conditions were included: near-threshold hit and miss, near-threshold correct rejection, and supraliminal hit and correct rejection. For each condition, the stimulus-induced activity (per voxel) was defined as the area under the curve (AUC) of 1.5 – 6.0 s (4 TRs) following the stimulus presentation. Note that conventional GLM analysis (often with zero baseline assumption) was not performed, because both pre-stimulus and post-stimulus effects were of interest in this experiment.

Statistical analysis in Exp1 and Exp2

For the pharmacological experiment, whole-brain voxel-wise one sample t tests (two-sided) against zero were performed at the group level for each activation map (e.g., tennis) and for each condition (e.g., Base1). For the psychological experiment, whole-brain voxel-wise paired sample t test (two-sided) was performed at the group level contrasting the stimulus-induced activity of hit (seen) versus miss (unseen) of a face in the near-threshold condition. All resulting z-maps were thresholded at the cluster level $\alpha < 0.05$. This was achieved using AFNI's upgraded function 3dttest++ with the '-Clustsim' option that simulates noise volume assuming the spatial auto-correlation function is given by a mixed-model rather than a Gaussian-shaped function (Cox et al., 2017).

Supplemental information

Anterior insula regulates brain network

transitions that gate conscious access

Zirui Huang, Vijay Tarnal, Phillip E. Vlisides, Ellen L. Janke, Amy M. McKinney, Paul Picton, George A. Mashour, and Anthony G. Hudetz
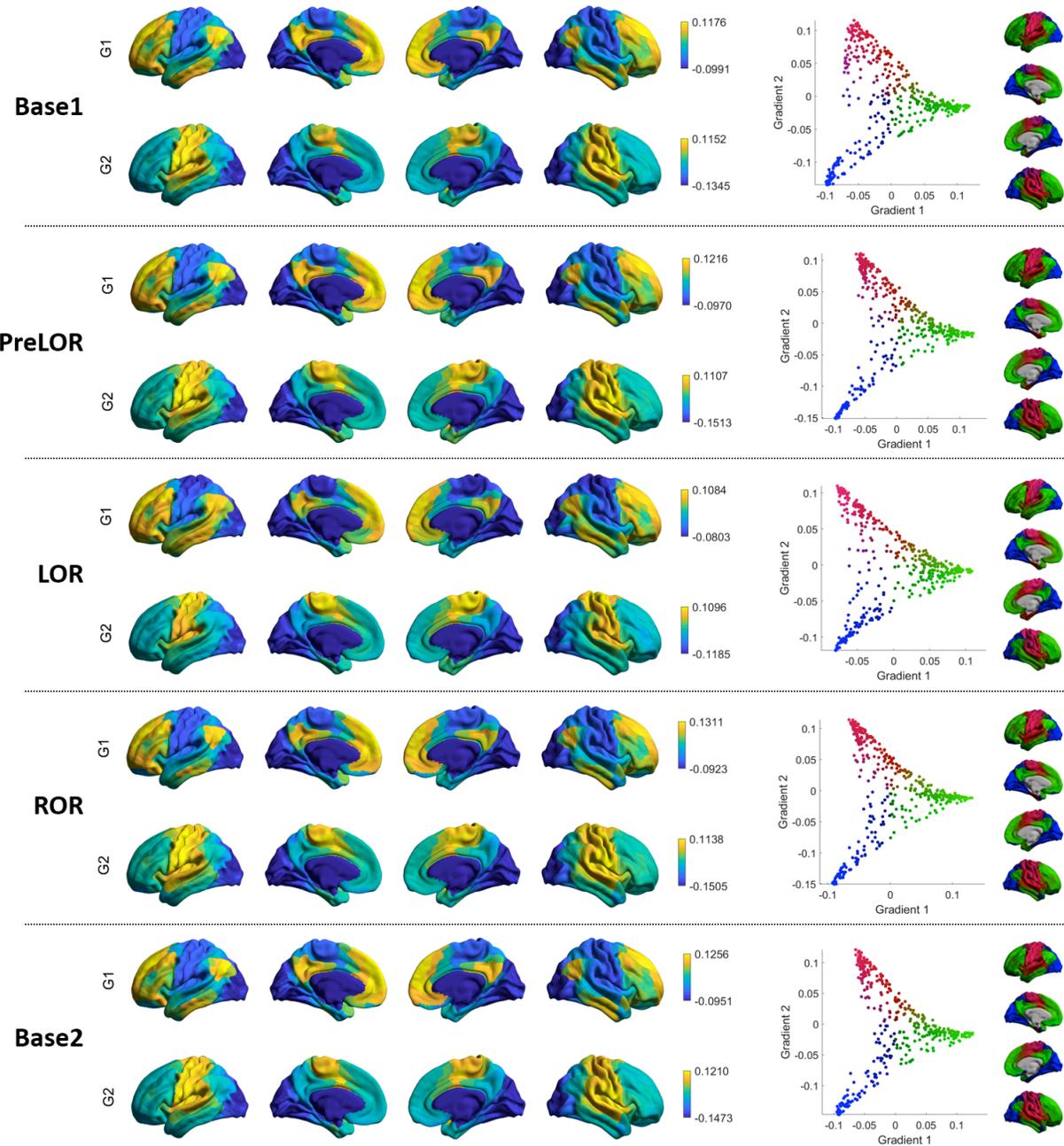
Figure S1. Cortical gradients across different conditions. Related to Figure 3. Left panel: topographic profiles of the first two gradients along the cortex. Right penal: the two gradients are projected into a 2-dimensional gradient space. The axes of the 2-dimensional space represent each gradient and separate distinct functional poles of cortical organization.
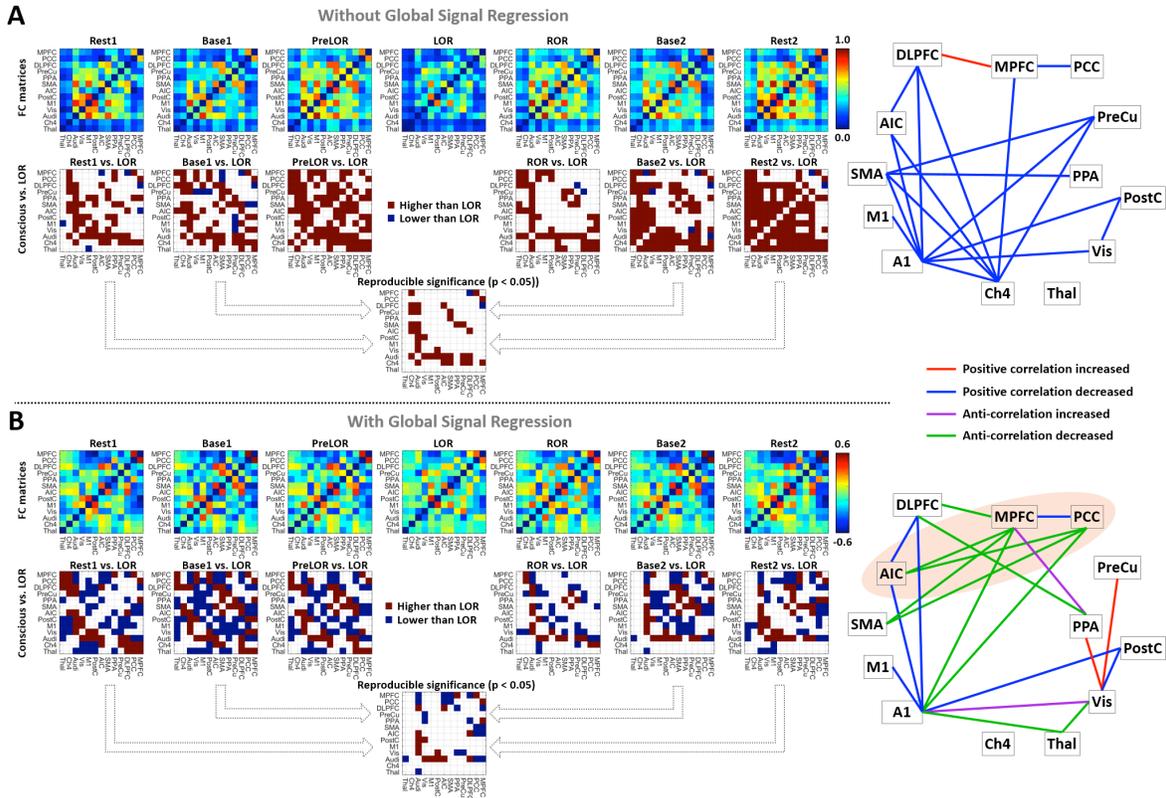
Figure S2. Functional connectivity among brain regions. Related to Figure 6. **(A)** Results without applying global signal regression (non-GSR). **(B)** Results after global signal regression (GSR). The time courses of fMRI signal change were extracted from the pre-defined ROIs for each condition, including Rest1 (n=26), Base1 (n=26), PreLOR (n=26), LOR (n=26), ROR (n=16), Base2 (n=25), and Rest2 (n=25). Qualitative estimation of functional connectivity was performed by calculating Pearson's correlation coefficient (Fisher's Z transformed) for each pair of ROIs. Individual pairwise functional connectivity matrices were generated for each condition. The reproducible differences across Rest1 vs. LOR, Base1 vs. LOR, Base2 vs. LOR, and Rest2 vs. LOR (two-sample paired sample t-tests; statistical significance at p<0.05) were considered as the genuine functional connectivity alterations during LOR (see connectivity matrix in lower, center position), with potential spurious functional connectivity driven by task-evoked activity accounted for. Schematics on the right summarize the effects of interest.
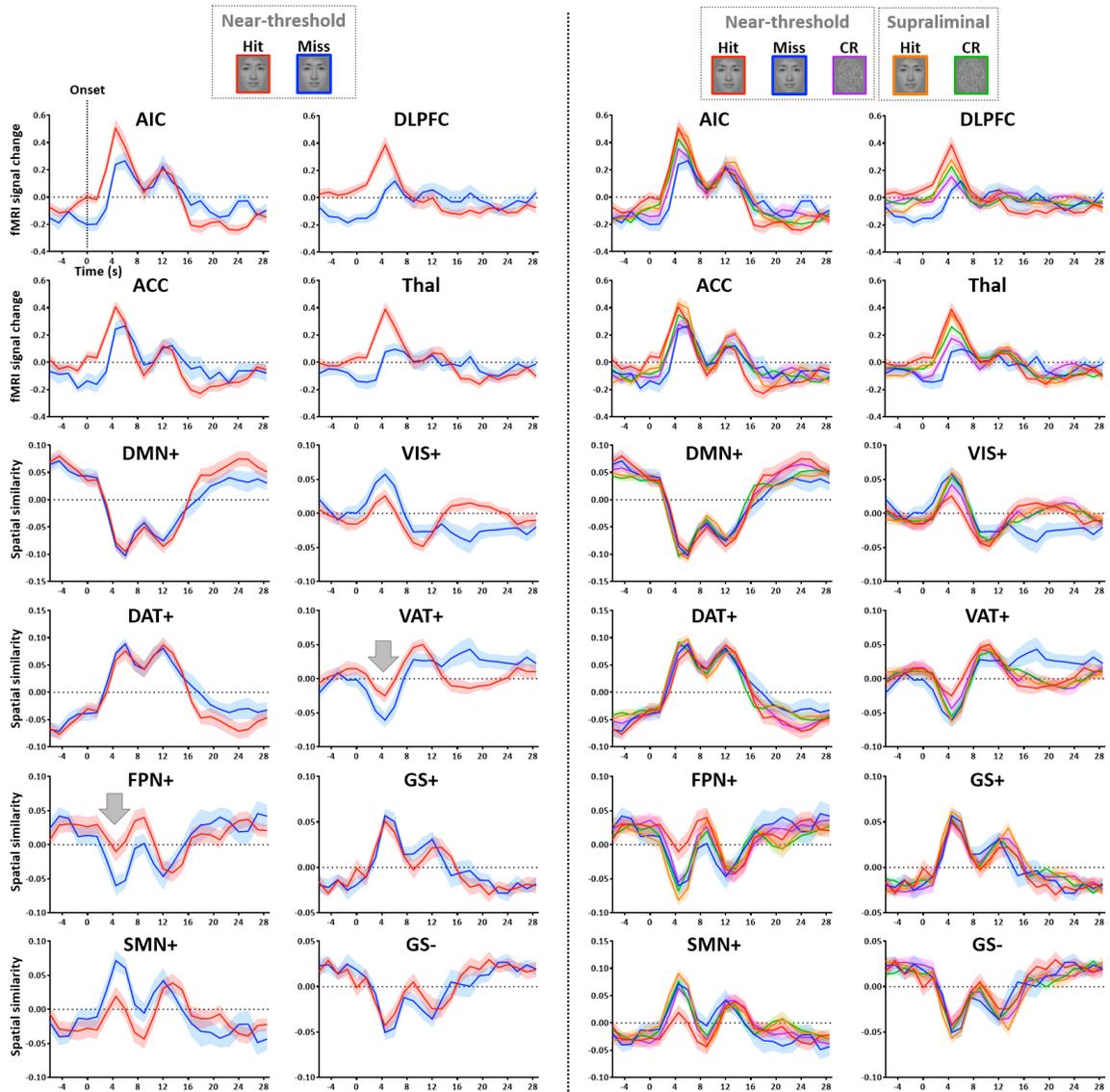
Figure S3. Time course of fMRI signal change (and spatial similarity of CAPs) for supraliminal and near-threshold conditions. Related to Figure 7. Five conditions were compared including near-threshold hit and miss, near-threshold correct rejection (CR), and supraliminal hit and correct rejection. AIC: anterior insular cortex; DLPFC: dorsal lateral prefrontal cortex; ACC: anterior cingulate cortex; Thal: thalamus. The co-activation patterns included the default-mode network (DMN+), dorsal attention network (DAT+), frontoparietal network (FPN+), somatomotor network (SMN+), visual network (VIS+), ventral attention network (VAT+), and global network of activation and deactivation (GN+ and GN-). Shaded areas indicate ±SEM across subjects (n=19).
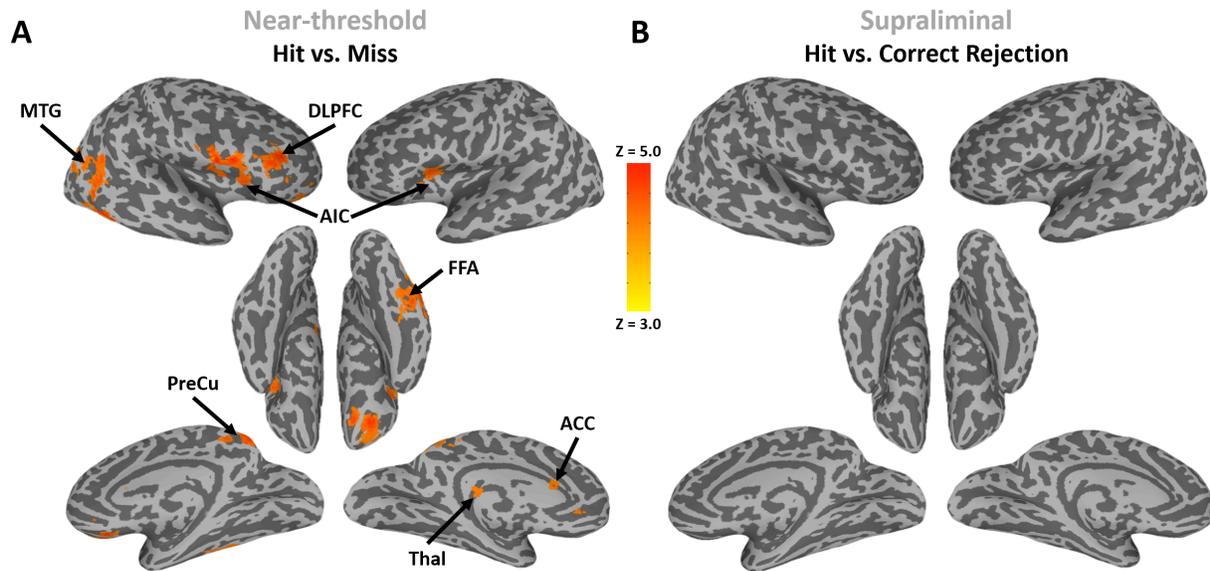
Figure S4. Whole brain contrast at the stimulus onset. Related to Figure 7. Group-level z-maps of fMRI signal intensity are shown for near-threshold hit vs. miss, and supraliminal hit vs. correct rejection (n=19). The z-maps were thresholded at cluster level α < 0.05. AIC: anterior insular cortex; DLPFC: dorsal lateral prefrontal cortex; ACC: anterior cingulate cortex; Thal: thalamus; MTG: medial temporal gyrus; PreCu: precuneus; FFA: fusiform face area.
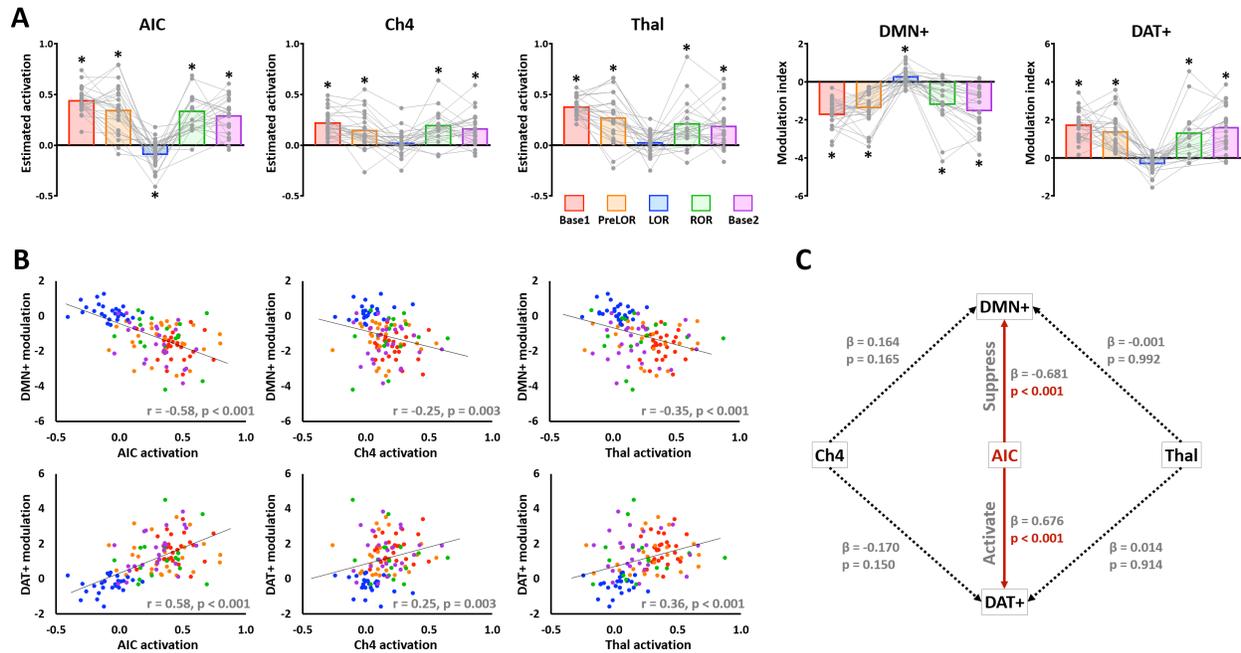
Figure S5. Relationship between AIC, Ch4, Thal and DMN-DAT switch. Related to Figure 6. **(A)** Instruction-evoked activation estimated from general linear model was plotted for AIC, Ch4 and Thal. Modulation indices were plotted for DMN+ and DAT+. Each grey dot represents an individual participant during Base1 (n=26), PreLOR (n=26), LOR (n=26), ROR (n=16), or Base2 (n=25) connected by grey lines across conditions. Bars represent the group averages for each condition. *indicates statistical significance (one sample t-tests against zero) at FDR-corrected alpha <0.05. **(B)** Pearson correlations (n=119) between the regional activation (AIC, Ch4 and Thal) and large-scale network modulation (DMN+ and DAT+). **(C)** Results from multivariate linear regression analysis (n=119) with the estimated activation in the AIC, Ch4 and Thal as independent variables, and the modulation indices in the DMN+ (and DAT+; analyzed separately) as dependent variables. Of the three independent variables tested, only AIC contributes with statistical significance to the DMN-DAT modulation.
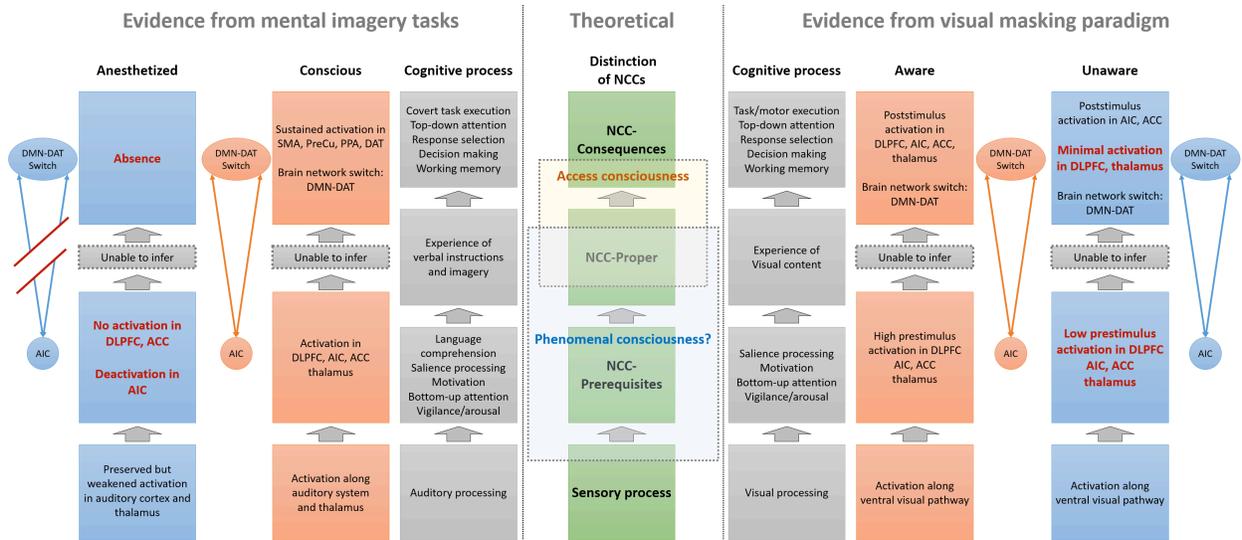
Figure S6. Schematic illustration of the rationale of experimental design and interpretation with respect to the terminology and theoretical aspect of consciousness. Related to Figure 4, Figure 5, Figure 6, and Figure 7. The middle column outlines different neural processes during conscious processing. The neural correlates of consciousness are theoretically dissociated from the prerequisites for and consequences of conscious processing (Aru et al., 2012). The left and right panels summarize the main results from the mental imagery and visual masking tasks. They are aligned to the presumed chains of cognitive processes and theoretical distinctions of NCCs.

Table S1. Summary of regions of interest. Related to Figure 3.

Anatomic structures	Abbreviations	BA	x [R]	y [P]	z [I]	z-value
Right Middle Temporal Gyrus	A1	21	59	-13	-7	7.90
Left Superior Temporal Gyrus	A1	21	-58	-19	2	7.66
Right Middle Frontal Gyrus	DLPFC	10	35	35	20	3.50
Left Middle Frontal Gyrus	DLPFC	10	-34	44	8	3.44
Left Insula	AIC	13	-31	20	2	4.72
Right Insula	AIC	13	32	17	2	4.28
Left Precentral Gyrus	M1	4	-36	-22	51	10.80
Right Medial Frontal Gyrus	MPFC	8	8	56	8	-4.07
Right Cingulate Gyrus	PCC	23	5	-49	26	-3.22
Left Precuneus	PPA	31	-13	-64	23	2.95
Right Posterior Cingulate	PPA	30	23	-58	17	3.07
Right Postcentral Gyrus	PostC	5	20	-43	65	-3.18
Left Postcentral Gyrus	PostC	5	-22	-37	68	-2.98
Left Precuneus	PreCu	7	-7	-70	41	4.77
Right Precuneus	PreCu	7	11	-70	41	4.25
Left Medial Frontal Gyrus	SMA	6	-4	2	50	6.84
Right Middle Occipital Gyrus	Vis	18	41	-82	2	-3.55
Left Inferior Occipital Gyrus	Vis	18	-37	-85	-4	-3.38

BA: Brodmann area; Talairach coordinates are provided for the peak voxel in a given region.